CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Epsilon Open Archive

Quantitative Genetics and Genomic Selection of Scots pine

Ainhoa Calleja Rodríguez
Faculty of Forest Sciences
Department of Forest Genetics and Plant Physiology
Umeå

Doctoral thesis
Swedish University of Agricultural Sciences
Umeå 2019

Acta Universitatis agriculturae Sueciae
2019:36

ISSN 1652-6880
ISBN (print version) 978-91-7760-390-0
ISBN (electronic version) 978-91-7760-391-7
© 2019 Ainhoa Calleja Rodríguez, Umeå
Print: Arkitektkopia, Umeå 2019

Quantitative Genetics and Genomic Selection of Scots pine

Abstract

The final objective of tree improvement programs is to increase the frequency of favourable alleles in a population, for the traits of interest within the breeding programs. To achieve this, it is crucial to decompose the phenotypic variance accurately into its genetic and environmental components in order to obtain a precise estimation of genetic parameters and to increase genetic gains. The overall aim of this thesis was to increase the accuracy of genetic parameter estimation by incorporating new quantitative genetics models to the analysis of multiple traits in multiple trials of Scots pine, and to develop a genomic selection protocol to accelerate genetic gain.

Factor analysis was incorporated to multivariate multi-environment analyses and it allowed to evaluate up to 19 traits simultaneously. As a result, precise patterns of genotype-by-environment interactions (G × E) were observed for tree vitality and height; moreover, it was possible to detect the main driver of the G × E: differences in temperature sum among sites.

Traditional quantitative trait loci (QTL) analysis of phenotypic data was compared with the detection of QTL with estimated breeding values (EBV) for the first time in a three generation pedigree and, as outcome, it was noticed that if a QTL was associated to a EBV and to a phenotypic trait, the proportion of variance explained by the QTL-EBV was higher than the QTL-phenotype. Additionally, several QTL were detected across several ages, which may make them suitable as candidates for early selection.

Genomic selection (GS) could aid to reduce the breeding cycle by shortening the periods of progeny field testing, and consequently increasing genetic gains per year. Genomic predictions, including additive and non-additive effects through different prediction models were compared with traditional pedigree-based models; it was seen an overestimation of genetic parameters for pedigree-based models, even larger when non-additive effects could not be discerned from additive and residual effects.

Prediction accuracies and abilities of the genomic models were sufficient to achieve higher selection efficiencies and responses per year varying between 50-90% by shortening 50% the breeding cycle. For the selection of the top 50 individuals, higher gains were estimated if non-additive effects are incorporated to the models (7 – 117%).

Keywords: Scots pine, genotype-by-environment, multiple variables, factor analysis, quantitative trait locus, genomic predictions, non-additive effects, Bayesian LASSO, Bayesian ridge regression, GBLUP.

Author's address: Ainhoa Calleja Rodríguez, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-901 83 Umeå , Sweden.
E-mail: Ainhoa.Calleja-Rodriguez@skogforsk.se

Dedication

A Güelita y Güelito

I am among those who think that science has a great beauty.
Marie Curie

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I **Ainhoa Calleja-Rodriguez***, Bengt Andersson Gull, Harry X. Wu, Tim J. Mullin, and Torgny Persson (2019). Genotype-by-environmental interactions and the dynamic relationship between tree vitality and height in Northern *Pinus sylvestris*. *Tree Genetics & Genomes*. doi: 10.1007/s11295-019-1343-8 (in press).

II **Ainhoa Calleja-Rodriguez**, Zitong Li, Henrik R. Hallingbäck, Mikko J. Sillanpää, Harry X. Wu, Sara Abrahamsson, and Maria Rosario García-Gil* (2019). Analysis of phenotypic- and Estimated Breeding Values (EBV) to dissect the genetic architecture of complex traits in a Scots pine three-generation pedigree design. *Journal of Theoretical Biology*, 462, pp. 283-292.

III **Ainhoa Calleja-Rodriguez**, Jin Pan, Tomas Funda, Zhi-Qiang Chen, John Baison, Fikret Isik, Sara Abrahamsson, and Harry X. Wu* (2019). Genomic prediction accuracies and abilities for growth and wood quality traits of Scots pine, using genotyping-by-sequencing (GBS) data (submitted to *G3: Genes, Genomes, Genetics*).

IV **Ainhoa Calleja-Rodriguez,** Zhi-Qiang Chen, Mari Suontama, Jin Pan and Harry X. Wu* (2019). Including non-additive genetic effects in genomic prediction of growth and wood quality traits in *Pinus sylvestris*. (Manuscript)

* Corresponding author.

The contribution of Ainhoa Calleja Rodríguez to the papers included in this thesis was as follows:

I She performed the data analysis, had a main role of writing the first draft of the manuscript and completed the paper in collaboration with co-authors.

II She contributed to the conceiving of the idea and experimental design, carried out the field work in the progeny trials, performed part of the data analysis and wrote the paper in collaboration with the co-authors.

III She was involved in the experimental design and performed part of the field sampling and laboratory work, she made the data analysis, had the main role of writing the first draft of the manuscript and completed the final manuscript in collaboration with co-authors.

IV She carried out part of the field sampling and laboratory work, she performed the data analysis and wrote the manuscript in collaboration with the co-authors.

List of tables

List of figures

Abbreviations

All abbreviations are explained when they first appear in the text.

1 Introduction

The objective of tree improvement programs is to increase the frequency of favourable alleles for traits of interest (that are generally quantitative traits) within the breeding programs. A breeding program involves different steps, including selection of elite trees, inter-mating or genetic testing of progenies derived from the inter-mating. Methodology of selection has advanced from phenotype to genotype of breeding values, and mating designs have also evolved from open-pollinated (OP) to control-pollinated with co-ancestry control. All these progresses are used in combination with quantitative genetics, among other things, to decompose the phenotypic variance into its components (additive, non-additive and environmental effects). The accuracy and reliability of such estimates depend on many factors, like, spatial variation, correlations among traits, different expressions of the traits in diverse environments, pedigree information and mating design, among others. The studies underlying this thesis used the traditional quantitative genetics and the more recent quantitative genomics methods, to improve the accuracy and estimations of genetic parameters of complex traits of interest within the Swedish Scots pine breeding program.

1.1 The species: Scots pine (*Pinus sylvestris* L.)

Scots pine (*Pinus sylvestris* L.) is one of the conifers most widely distributed throughout the whole Eurasia, ranging between latitudes 37°N – 70°N, i.e., from southern Spain to northern Scandinavia (Figure 1); with a longitudinal range that varies from 8°W (Portugal) to 141°E (east of Russia). The species can grow in very diverse soil conditions and at different elevations from the sea level up to 2700m (Boratyński, 1991).

The temperature sum (Tsum) or temperature climate, i.e., the summation of all daily mean values exceeding the threshold value of +5°C (Bärring *et al.*,

2017), is crucial for the flowering to begin, as well as, for shoot elongation, requiring lower Tsum in colder regions than in southern regions (Mátyás *et al.*, 2004).

From the commercial point of view it is a particularly important species in the Nordic countries, where its major uses are, among others, construction wood, furniture, fibre- and chip-boards, wood pulp and paper (Krakau *et al.*, 2013). In Sweden, it is the second foremost species, representing the 39% of stand volume production (The Swedish National Forest Inventory, 2015). According to Krakau *et al.* (2013) only 1–2% of stands in Sweden are set by direct seeding, 20–25% of stands are established by natural regenerations whereas the majority of them are planted; besides, the 80% of those plants have their origin in seeds from improved seed orchards.

Figure 1. Distribution map of Scots pine (Caudullo *et al.*, 2018) [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)].

1.2 Scots pine breeding in Sweden

Forest tree breeding, combined with forest management, has become an essential part of the forestry operations by supplying improved individuals for traits of particular interest to the industry (Pâques, 2013). It involves multiple steps, such as, inter-mating, genetic testing and selections, with the final aim of increase the frequency of favourable alleles of those important traits in breeding populations (Grattapaglia, 2014). The final goal of each breeding program is to produce improved trees in an economically efficient way, by maximizing the genetic gain (i.e., a higher frequency of favourable alleles than the earlier population) per unit time at the lowest possible cost (White *et al.*, 2007).

Possibly due to its commercial importance, Scots pine is considered the most studied tree species from the point of view of provenance research and its improvement started already around 100 years ago (in 1907) with the

establishment of an international IUFRO provenance study with individuals from different climatic regions (Mullin *et al.*, 2011).

The Swedish Scots pine tree improvement program started between the 1950-1960's, by selecting around 1300 plus-trees from natural stands based on their superior phenotypes and grafting them into seed orchards or clonal archives, known as the first round of seed orchards (Andersson *et al.*, 2003). Those trees became the parents of the progeny trees which started to be tested in the field in the 1970's – 1980's and are part of the actual Scots pine breeding program (Mullin *et al.*, 2011). In the early 1980's about 4700 plus-trees were selected, comprising the second round of seed orchards. Roughly, those 6000 plus-trees conformed the base material used to establish the founder populations for the long-term Scots pine breeding program in Sweden (Wilhelmsson & Andersson, 1993). Approximately a 6% and 10% of gain in production at full rotation was obtained from the first and second rounds of seed orchards, respectively. However, survival rates were not improved; for the third round of seed orchards from tested progeny, it is expected to obtain gains among 23–27% in volume production and 5–15% in survival rates (Andersson *et al.*, 2007; Rosvall *et al.*, 2001).

Around 1500 selected parent trees conform a meta-population that was divided into 24 breeding populations, each containing between 50 to 70 individuals. Long term breeding is managed for each of the breeding populations that are distributed overlapping each other according to different adaptation targets based on photoperiod (latitude) and Tsum (Figure 2), covering an area larger than the actual Tsum range of Sweden (Danell *et al.*, 1993).

The first generation (F_1) progeny trials are formed by full-sib (FS) families (generated by partial diallel mating design) or half-sibs (generated by poly-crosses or OP seeds from plus-tree stands). From those tested trees a double-pair mating design and, if possible, positive assortative mating are used to generate the second generation (F_2) of progeny trials, to test families in four different sites and perform essentially within-family selection for the next generation (Krakau *et al.*, 2013; Ruotsalainen & Persson, 2013).

The Swedish Scots pine breeding program combines a series of different selection strategies, classified here into two groups according to the cycle lengths (Rosvall, 2011):

- The first group of strategies consist of both forward and backward selections, requiring natural or stimulated flowering to produce new generations of progeny tests, thereby the breeding cycle length takes up to 36 years.

- The cycle length of the second group takes up to 21 years, and among- and within-family selections are made forward, based on FS or poly-cross family tests.

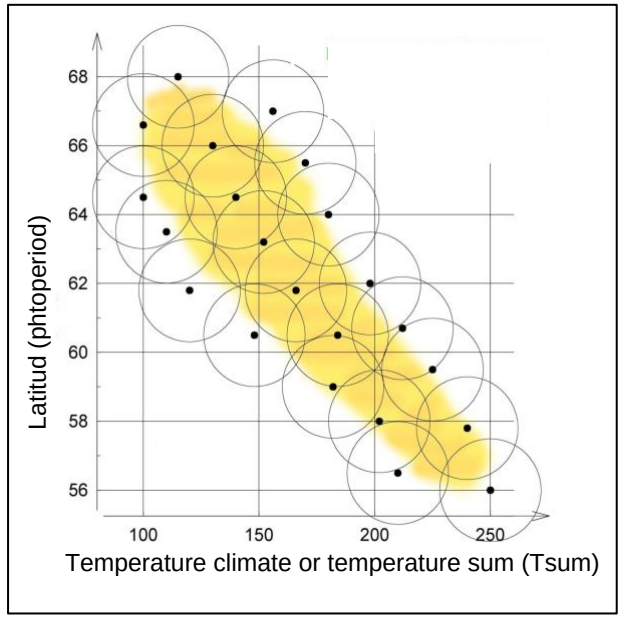

Figure 2. Scots pine breeding populations.

All strategies are used to achieve the general breeding objectives which are essentially to 1) improve traits of economic value (vitality, growth, wood density, stem-wood quality, survival and resistance against biotic and abiotic stresses), 2) prepare for new climatic conditions derived from the climate change (plasticity/adaptability), and 3) assure the adequate genetic diversity in the breeding populations (Andersson *et al.*, 2011; Danell *et al.*, 1993).

1.3 Quantitative genetics

Most of the traits of interest within the breeding programs are complex traits also known as polygenic or quantitative traits, as they are controlled by many loci each of which has a small effect on the phenotypic expression of the trait (Goddard & Hayes, 2009). Complex traits differ from Mendelian traits in that the former ones must be studied at population levels (i.e., large numbers of related individuals), because in addition to their polygenic attributes, their

phenotypic expression is also affected by environmental effects (White *et al.*, 2007; Falconer & Mackay, 1996).

Quantitative genetics study quantitative traits using diverse statistical methods that essentially divide the previously quantified phenotypic variation of populations into their genetic and environmental components, to predict the genetic values of each individual or family studied (Lynch & Walsh, 1998).

1.3.1 Genetic parameter estimation

The major goal of the breeding studies is to estimate the proportion of the genetic variance that is transmitted to the offspring, i.e., the additive variance or breeding value; and, to estimate heritabilities of the traits, i.e., the ratio between the additive variance and the phenotypic variance (White *et al.*, 2007; Falconer & Mackay, 1996). Such estimations can only be performed through field tests with different family structures and mating designs, that facilitate the decomposition of the phenotypic variance into its components (genetic and environmental variances). Moreover, genetic gains rely on the accuracy and reliability of the estimates since they can be biased by different factors such as, spatial variation, interactions between different genotypes and environment, pleiotropy among different traits, etc., as it is described below.

1.3.2 Spatial analysis

Spatial analysis can improve the estimations of genetic parameters by discerning the spatial and non-spatial components of the environmental variation and therefore reducing error variance (Chen *et al.*, 2018c; White *et al.*, 2007). Spatial analysis can detect global trend (gradient), local trend (patchy) and extraneous (nugget) variations across large heterogeneous agricultural and forest field trials used in breeding programs, with the consequent reduction of the residual errors and improvement in the accuracy of genetic parameter estimations (Dutkowski *et al.*, 2002; Cullis *et al.*, 1998; Gilmour, 1997). Different methods of spatial analysis have been studied and applied like post-blocking (Gezan *et al.*, 2006; Ericsson, 1997) or kriging (Zas, 2006). However, a method that fits both design features and the spatial component as first-order separable autoregressive model of residuals, had shown less bias in the estimation of genetic parameters (Dutkowski *et al.*, 2006; Dutkowski *et al.*, 2002; Gilmour, 1997). The latter method has aided to increase the accuracy of breeding values estimates by reducing residual variation in different tree species, like at Cappa *et al.* (2017); Silva *et al.* (2013) in *Eucalyptus grandis*, Bian *et al.* (2017) in Chinese fir (*Cunninghammia lanceolata* (Lamb.) Hook), Chen *et al.* (2018c) in Norway

spruce (*Picea abies* (L.) Karst.) or Resende *et al.* (2016) in eucalyptus hybrids (*Eucalyptus grandis* × *E. uropphylla*) or Ivkovic *et al.* (2015) in radiata pine (*Pinus radiata* D. Don); therefore, it was the method used to perform spatial analysis in all Scots pine trials analysed in this thesis (Papers I – IV).

1.4 Multi-environment trials (MET) and genotype-by-environment (G × E) interactions

With the objective to obtain unbiased variance components estimates, field tests must be established at different locations (White *et al.*, 2007), as the estimates in one single site can be overestimated, due to the different expression of the genetic variance in different sites (Burdon *et al.*, 2019; Li *et al.*, 2017; Burdon, 1977). Multi-environment trials (MET) allow to select individuals that perform better across multiple environments, as well as to detect the degree and pattern of genotype-by-environment interactions (G × E interactions or G × E); besides, in the case that some unexpected event occur, MET aid reduce the risk of losing genetic material (Isik *et al.*, 2017).

Genetic correlations between the same trait in different environments, known as Type-B genetic correlations (Burdon, 1977), are commonly used to evaluate the degree and pattern of G × E (Burdon *et al.*, 2019; Chen *et al.*, 2017; Berlin *et al.*, 2015; Baltunis *et al.*, 2010; Hannrup *et al.*, 2008). Type-B genetic correlations and patterns of G × E were studied in Papers I and II. Different traits can be correlated as well, and the main cause could be pleiotropy, i.e., a gene that affects the expression of more than one trait (He & Zhang, 2006). Correlations among traits are important since they can affect the response to selection of correlated traits i.e., indirect selection (White *et al.*, 2007).

1.4.1 Univariate and multivariate MET analysis

If multiple traits are simultaneously analysed (multivariate analysis), more complex patterns of G × E can be detected such as, Type-A genetic correlations, *i.e.*, correlations between different traits within the same environment, and Type-AB genetic correlations, i.e., correlations between different traits expressed among environments (Li *et al.*, 2017; Dutkowski *et al.*, 2016). Multivariate (MV) analysis are preferable over univariate (UV) analysis because they can increase the accuracy of breeding values and heritabilities by taking advantage of the possible hidden covariances and correlations existing among traits (Isik *et al.*, 2017; Mathew *et al.*, 2016). Simulation studies have shown the selection response is larger with MV MET (Bauer & Leon, 2008) than UV MET analyses, especially for trees located in trials in which they are exposed to non-random

mortality, because UV MET analysis cannot take into account the factors to which the selection process is related, like survival in Scots pine trials located in northern areas (Persson & Andersson, 2004). Large mortality rates of Scots pine occur in harsh areas of Northern Sweden; previous studies at early ages showed evidence that there is a negative association between tree vitality and height, but suggested to perform more studies at adult ages to confirm this theory (Persson, 2006). Hence, within the current thesis we studied the association between tree height and survival in harsh and mild areas in four different populations of the Scots pine breeding program, at both early and mature ages (Paper I), by implementing UV and MV MET analysis with real assessments, to estimate Type-A, Type-B and Type-AB genetic correlations, as well, as patterns of G × E, among tree vitality (i.e., a measure of tree health) and height.

1.4.2 Factor analysis

Diverse methodologies have been regularly used to analyse G × E and their patterns, such as, analysis of variance (ANOVA), principal component analysis or linear regression (Meyer, 2009; Freeman, 1973), still, they are not the most adequate methods to explore unbalance MET data, and in addition, they cannot discern G × E patterns; therefore, to achieve more accurate estimates of genetic parameters and patters of G × E, more complex methods that account for the heterogeneity of variance-covariance components across different environments are needed (Kelly *et al.*, 2007; Cullis *et al.*, 1998). The unstructured (US) variance-covariance matrix in mixed models is the general model that accounts for site specific variances and heterogeneous covariances among sites; nevertheless, with large number of environments and traits, US models are computational restrictive and can end up with convergence problems, which make their real applicability very limited (Isik *et al.*, 2017).

Factor analytic (FA) structures, also account for site specific variances and heterogeneous covariances among sites and are considered a good approximation to US models requiring less parameters (Isik *et al.*, 2017; Smith *et al.*, 2001). FA models are widely applied in crops (Smith *et al.*, 2015; Beeck *et al.*, 2010; Kelly *et al.*, 2009; Kelly *et al.*, 2007; Smith *et al.*, 2001), and recently have been applied in several forest tree species (Walker *et al.*, 2019; Smith & Cullis, 2018; Ukrainetz *et al.*, 2018; Chen *et al.*, 2017; Gezan *et al.*, 2017; Ivkovic *et al.*, 2015; Cullis *et al.*, 2014; Ogut *et al.*, 2014; Hardner *et al.*, 2010). In the studies carried out in this thesis (Papers I – II), it has been studied the utility of FA for UV and MV MET analysis, for the first time in Scots pine MET data within the Swedish breeding program.

1.5 Genomic selection

1.5.1 Association- and quantitative trait loci- mapping

The development of marker technologies and the consequent discovering of large number of polymorphic markers, can help to dissect the genetic architecture of complex traits. Quantitative-trait-loci (QTL) mapping and association mapping (AM), were developed to better understand the architecture of complex traits. QTL mapping aims to identify the regions where genes have influence in the quantitative trait, and can help to approximately locate, estimate the number, size and effect of the genes affecting the trait; whereas, AM can detect the genes that affect the trait more precisely, and the mutations that have effect on the phenotypic differences of the trait among individuals (White *et al.*, 2007).

One of the most important challenges for forest tree breeding programs is the long term of breeding cycles. Marker-aided-selection (MAS) was thought to be the instrument to short breeding cycles and better understand complex traits (Neale & Williams, 1991). Generally, MAS consist of selecting individuals (normally biparental populations) with QTL-associated markers that have major effects, and use those QTL to make decisions in breeding (Grattapaglia, 2014; Neale & Williams, 1991); however, MAS was not really implemented in tree breeding because, among others, 1) it is very limited by the genetic background, i.e., cannot be applied in different families or populations than those involved in the QTL study; 2) G × E or QTL-by-environment (QTL × E) interactions, i.e., different expression of QTL across environments; 3) low linkage disequilibrium (LD) at population levels in tree breeding, 4) the polygenic nature of traits (White *et al.*, 2007; Strauss *et al.*, 1992). Additionally, a challenge of AM is the detection of false positives and overestimation of QTL effects (Isik, 2014), but this can be address aided by new methodologies recently implemented (Li *et al.*, 2017). With the aim to dissect the genetic architecture of complex traits (adaptive and growth traits) a de-biased LASSO method was implemented to detect QTL in a three generation pedigree of Scots pine (Paper II).

1.5.2 From MAS to Genomic Selection

The limitations of traditional MAS defined in the above section, were overcome by the genomic selection (GS) methodology proposed by Meuwissen *et al.* (2001), that uses all available genome-wide markers simultaneously to predict genomic estimated breeding values (GEBV) with the help of additive genetic models, and/or the total genomic estimated genetic values (GEGV) when non-

additive models are used. A notable difference between traditional MAS and GS is that the latter one does not need to detect the marker-trait associations or QTL prior to selection (Isik, 2014; Grattapaglia & Resende, 2011).

The implementation of GS in breeding programs essentially requires two phases; 1) developing prediction models in a population that is phenotyped and genotyped, known as training population or training set (TS); and, 2) performing cross validation of the prediction models in selection candidates genetically related with the TS, known as validation population or set (VS), that are only genotyped and for which GEBV are predicted (Meuwissen *et al.*, 2016; Goddard & Hayes, 2009).

It is expected to obtain higher genetic gains through GS mainly due to a reduction of the breeding cycle lengths by shortening field test time through early selections obtained by GS predictions, as it has been shown already in animal and crop breeding (Crossa *et al.*, 2017; Meuwissen *et al.*, 2016). Higher selection intensities would be possible, since more progenies could be genotyped than those that nowadays are established in field trials (Grattapaglia *et al.*, 2018; Crossa *et al.*, 2017).

Several GS prediction studies have been recently published in forest trees (Chen *et al.*, 2018b; Lenz *et al.*, 2017; Ratcliffe *et al.*, 2017; Isik *et al.*, 2016; Beaulieu *et al.*, 2014; Resende *et al.*, 2012a; Resende *et al.*, 2012c). This thesis comprises the first attempt (or proof-of-concept) to study GS in the Swedish Scots pine breeding program (Papers III – IV).

1.5.3 Additive and non-additive effects through GS

Phenotypic variance, in addition to additive and environmental effects, is also influenced by non-additive effects, which are not directly transmitted from parent to offspring (White *et al.*, 2007). Non-additive effects can help to obtain more reliable and unbiased estimations of the additive effect; for example, if only a small proportion of the phenotypic variance is due to additive effects, the remaining will be due to environmental and non-additive effects, and it is important to discern among them to get unbiased estimates of narrow-sense heritabilities and breeding values (Walsh & Lynch, 2018; Lynch & Walsh, 1998; Falconer & Mackay, 1996).

Additional advantages of GS are that it does not rely on the pedigree, which is susceptible to errors; and, that GS can predict GEGV, i.e., additive and non-additive effects (dominance and epistasis) without establishing complicated and expensive mating designs, needed when phenotypic and pedigree data are used to predict estimated breeding values (EBV) or genetic values. In particular, with the traditional pedigree-based genetic analysis, the genetic variance components

can only be discerned when FS families (additive and dominance components) or FS families with replicated clone structures (additive, dominance and epistatic components) are available. That might be the reason why only a few forest species have been studied to evaluate the non-additive variation based on pedigree (Baltunis & Brawner, 2010; Isik *et al.*, 2004; Isik *et al.*, 2003). El-Dien *et al.* (2016) and El-Dien *et al.* (2018) were able to estimate GEGV aided by genomic prediction methods in an OP population. In the studies underlying this thesis, both additive (Paper III – IV) and non-additive effects (Paper IV) with pedigree and genomic information have been estimated and compared.

1.5.4 Prediction accuracies

According to Hayes *et al.* (2009), several factors can affect the accuracy and prediction ability (PA) of genomic predictions: 1) the level of extent of LD between markers and QTL, which is dependent on the effective population size (N_e) and the number of markers used; 2) the number of individuals in the TS from which the marker effects are estimated; 3) the heritability of the trait under study; 4) the marker density (distribution of QTL effects). Moreover, predictions depend also on G × E, age-age correlations and the statistical method used to perform predictions (Grattapaglia *et al.*, 2018; Isik, 2014).

Different statistical methods are available to estimate GEBV and differ with respect to the genetic architecture of the trait or QTL effects. GBLUP (genomic best linear unbiased prediction) and RR-BLUP (ridge-regression BLUP) assume that QTL allelic effects are normally distributed and all have similar contribution to the genetic variance (Isik *et al.*, 2017; Endelman, 2011), whereas Bayesian approaches presume a prior gamma or exponential distribution of QTL allelic effects, thus that the variance at each locus can vary (Isik *et al.*, 2016; Resende *et al.*, 2012c; Meuwissen *et al.*, 2001). In the literature, several methods have been tested in forest trees with similar results (Chen *et al.*, 2018b; Suontama *et al.*, 2018; Tan *et al.*, 2017; Thistlethwaite *et al.*, 2017; Bartholome *et al.*, 2016b; Ratcliffe *et al.*, 2015).

Genomic predictions can be improved through the inclusion of all components of the genetic variance, i.e., additive, dominance and epistatic effects, as several studies in forest tree species have already demonstrated (Chen *et al.*, 2018a; Tan *et al.*, 2018; Bouvet *et al.*, 2016; de Almeida *et al.*, 2016; Munoz *et al.*, 2014). Accuracies of genomic predictions and PA have been studied within this thesis; several BLUP and Bayesian genomic prediction models have been tested (Paper III), and predictions with only additive (Papers III – IV) and with both additive and non-additive effects (Paper IV) have been compared.

1.5.5 Next generation sequencing (NGS)

The change from MAS to GS, was only possible through the development of next generation sequencing technologies (NGS) that allowed the detection of thousands of single nucleotide polymorphism (SNP) markers in a more cost-effective way, such as, SNP arrays or exome probe panels (Grattapaglia *et al.*, 2018). For forest tree species, such as Scots pine, genotyping-by-sequencing (GBS) is an attractive and alternative genotyping method since the species has large genome, absence of reference ones and for which SNP arrays or exome panels are not yet developed (Dodds *et al.*, 2015; Chen *et al.*, 2013; Elshire *et al.*, 2011). GBS has been successfully used in genomic predictions of animal, crop and tree breeding (El-Dien *et al.*, 2018; El-Dien *et al.*, 2015; Gorjanc *et al.*, 2015; Ratcliffe *et al.*, 2015; Crossa *et al.*, 2013; Poland *et al.*, 2012; Poland & Rife, 2012). GBS data were used to perform genomic predictions in Scots pine (Papers III – IV).

2 Objectives

The overall aim of the studies underlying this thesis is to estimate genetic parameters and breeding values for vitality, growth and wood quality traits of Scots pine in Sweden, by incorporating factor analytic and genomic prediction methods, as well as, delving into the dissection of complex traits. The following specific objectives were addressed:

- To explore whether there are G × E patterns for survival, growth and adaptive traits in Scots pine trials of Northern Sweden (Papers I – II); to test if genetic correlations among tree vitality and height vary depending on the harshness of the sites (Paper I); and, to study if more accurate estimations of genetic parameters and more complex patters of G × E, are obtained with MV MET analysis compared with UV MET analysis (Paper I).

- To dissect the genetic architecture of adaptive and growth traits, by using phenotypic data and EBV, from a three generation MET data, (Paper II).

- To predict GEBV and GEGV with GBS data for growth and wood quality traits in Scots pine (Papers III – IV); to evaluate predictive abilities and accuracies based on different genomic prediction models and to examine whether the size of TS and VS, and the number of SNPs have any effect on the prediction abilities and accuracies (Paper III).

- To incorporate non-additive effects to genomic and pedigree prediction models and assess their prediction abilities for growth and wood quality traits (Paper IV); to estimate genetic parameters considering non-additive effects, and to compare the precision of the EBV, GEBV and GEGV (Paper IV).

- To compare the relative genomic selection efficiency (Paper III) and the expected response of genomic selection (Paper IV), with the traditional phenotypic pedigree-based selection.

3 Materials and methods

3.1 Scots pine field trials

All trials addressed in this thesis belong to the Swedish Scots pine breeding program and are located in middle and northern Sweden (Figure 3).

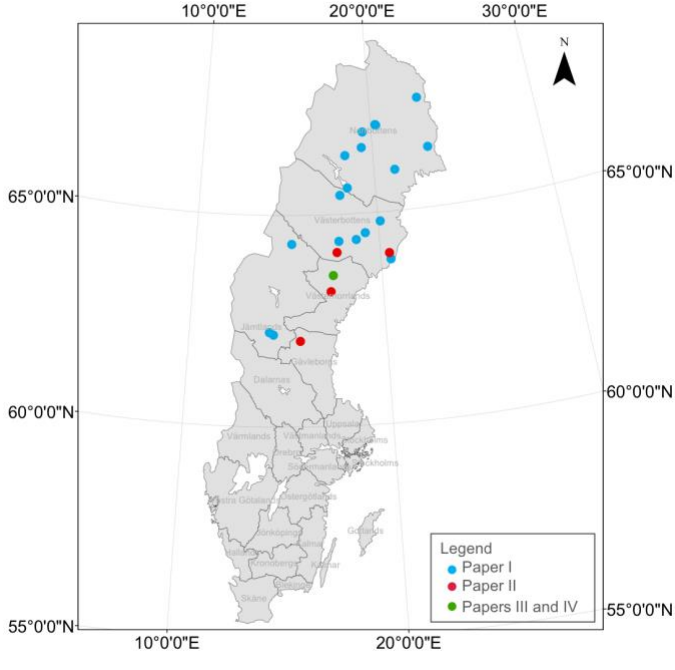

Figure 3. Location of all trials

The trials belong to five different breeding populations within the Scots pine breeding program and were all planted in randomized single tree plot designs.

Four unrelated populations (Pop 1, 3, 5 and 6) or trial series of, in total, 20 progeny trials were used in Paper I; every trial includes between 288 and 360 OP families originated from plus tree selection. Each trial series contain five field trials and almost all the OP families were represented in all trials within each series (Figure 4). Trial series 1 (Pop 1) was planted between 1983 – 1986, whereas trial series 2 (Pop 6) and 3 (Pop 5) were respectively planted in 1990 and 1991. The trial series 4 (Pop 3) was planted between 1993 – 1994. The total amount of individuals planted per trial varied between 6061 – 8116 at Pop 1 and 5, and among 3027 – 4207 for Pop 6 and 3.

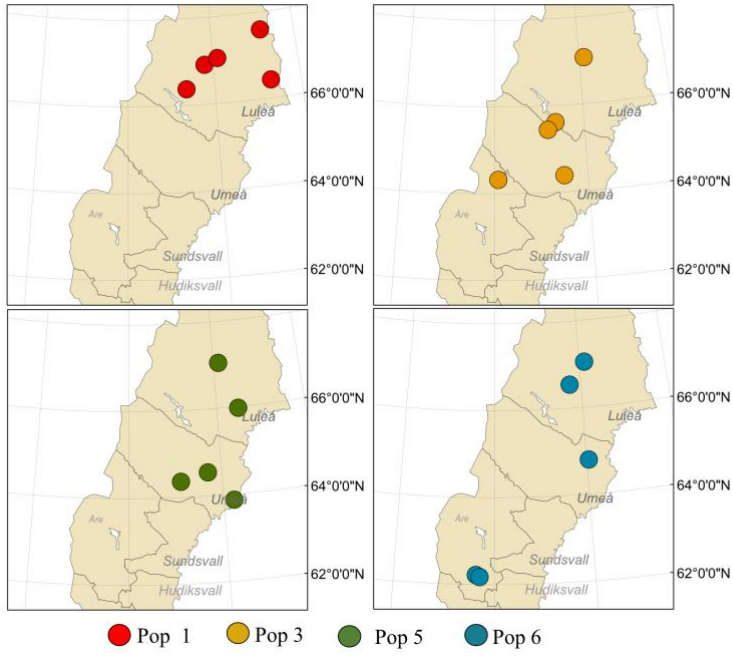

Figure 4. Location of the open-pollinated progeny trials within each breeding population, used in Paper I.

From breeding Pop 11, two plus-trees, AC3065 and Y3088 (F_0-generation), were selected for their superior performance in field evaluations (Figure 5). AC3065 was selected as female and artificially pollinated with Y3088. In 1988, 1000 seedlings originated from the control-cross were established in the field (trial F485), thus creating one of the largest FS families planted in one field trial in Sweden (F_1-generation). A total of 455 FS individuals were selected for Paper

II. 360 trees, out of the 455 FS individuals, produced wind pollinated cones in 2006, resulting in OP seedlings (F_2-generation). No pollen production was observed therefore it was assumed that pollination occurred from external resources without inbreeding. The OP seedlings were grown at the Skogforsk nursery in Sävar and, in 2008 the OP individuals were planted in three different F_2 field trials (F723, F725 and F726), and were also used in Paper II.

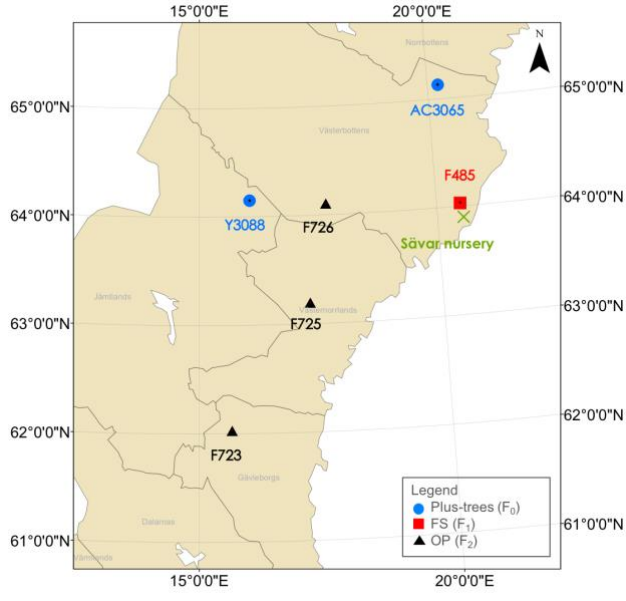

Figure 5. Location of the plus-trees and, full-sib (FS) and open-pollinated (OP) field trials studied in Paper II.

For Papers III – IV, 694 progeny trees from 183 FS families were selected from trial F261 which belongs to breeding Pop 11. The field trial (F261, Grundtjärn) was established in 1971 and is located at latitude 63.55°N and longitude 17.42°E (green circle in Figure 3); it is composed of 184 families and 7240 trees generated from a partial diallel mating design of 40 plus-trees and five reference seed-lots.

3.2 Traits studied

All traits addressed in this thesis and their age of assessment as well as the study and field trial in which they were measured are summarized in Table 1, and are defined underneath.

Table 1. *Traits assessed and their acronyms, the age of assessment and trial and paper for which they were assessed.*

Trait	Acronym	Age assessment	Paper / Trial
Height	Ht	10, 20	I / all
		3	II / F_2 (F726)
		8	II / F_2 (all)
		8 – 18	II / F_1 (FS)
		10, 30	III, IV
Diameter at breast height	DBH	8	II / F_1 F_2
		11, 16, 17	II / F_1
		30, 36	III, IV
Branch diameter	BrD	8, 11, 16	II / F_1
Branch quality	BrQ	8	II / F_2
Branch angle	BrA	8	II / F_1 F_2
		11, 16	II / F_1
Microfibril angle	MFA	40	III, IV
Static modulus of elasticity	MOEs	40	III, IV
Wood mean density	DEN	40	III, IV
Dynamic modulus of elasticity	MOEd	40	III, IV
Tree vitality	Vt	10, 20	I / all
		1, 8	II / F_2
		3	II / F_2 (F726)
Cold hardiness	Ct	3	II / F_2 (F726)
		4, 9, 11, 16	II / F_1
Flower production	FP	18	II / F_1
Number of cones per tree	CO	18	II / F_1
Number of seeds per cone	nSC	.	II / F_1
Weight of 1000 seeds	W	.	II / F_1

3.2.1 Height

Height (Ht) was measured in all trials studied in this thesis (Papers I – IV) although it was assessed at different ages depending on the study, as it is explained below.

In Papers I, III and IV, trees were assessed at two different ages; in Paper I, the first age of assessment varied between 9 – 13 years, and the second age of assessment between 18 – 22 years, depending on the trial, however within each series, trials were measured with a maximum difference of two years; for simplicity, we denote those assessments as Ht10 and Ht20 (Ht1 and Ht2, in Paper I), to refer to the first and second assessments respectively. In Papers III and IV,

all individuals in the trial were assessed at ages 10 and 30, i.e., Ht10 and Ht30, respectively.

In Paper II, Ht measurements were assessed at different ages depending on the F_1 and F_2 trials. The F_1 trial (F485, in Figure 5) was measured annually between ages 8 – 18, whereas F_2 trials were measured at age eight; additionally, the F_2 F726 trial (Figure 5) was measured at age three in a former study (Abrahamsson *et al.*, 2012), and was chosen due to its highest survival rates compared with the two remaining F_2 trials.

3.2.2 Diameter at breast height

Diameter at breast height (DBH) was measured at age eight in all F_2 trials in Paper II; at ages 11, 16 and 17 in the FS trial from Paper II, and at ages 30 and 36 in the trial used in Papers III – IV.

3.2.3 Tree vitality (survival ability)

Mortality of Scots pine in harsh areas is higher than in milder ones and normally is caused by diverse events as a result of accumulated damages over several years (Stefansson & Sinko, 1967; Eiche, 1966); mortality occurs predominantly during the first 12–16 years after planting and decreases considerably after 20 years, approximately (Persson & Ståhl, 1993). Tree vitality (Vt), i.e., a measure of survival ability, was scored according to Persson & Andersson (2003) in four categorical classes: healthy, slightly damaged, severely damaged but alive, and dead (or missing).

In Paper I, Vt was assessed two times at two different ages, which are the same as described before for Ht; for simplicity, we denoted Vt10 and Vt20 for all trials (Vt1 and Vt2, in Paper I), to refer to the first and second assessments respectively. Vt was scored at the age of eight in all OP trials used for Paper II, and additionally at age three in the F726 trial.

3.2.4 Wood quality traits

Brach properties

Branch angle (BrA), branch quality (BrQ) and branch diameter (BrD) are routinely measured within the Scots pine breeding program, as they are an indicator of wood quality and biomass (Mäkinen & Colin, 1998). BrA, BrQ and BrD were studied in Paper II. BrD was only measured in the individuals from the FS trial, at ages 8, 11, 16; BrA was assessed at the same ages as BrD for the

FS trial and at age 8 in the OP field trials, where BrQ was assessed too (at age eight).

Silviscan assessments

Increment cores at breast height were extracted for 694 individuals from trial F261 and processed by Silviscan (Innventia AB, Stockholm, Sweden). In Papers III and IV, three solid-wood quality traits were addressed from the Silviscan analysis: wood density (DEN), microfibril angle (MFA) and stiffness (expressed in terms of static modulus of elasticity: MOEs).

Acoustic velocity

The dynamic modulus of elasticity (MOEd) predicted by Hitman ST300 (Fibergen, Christchurch, New Zealand) and further describe in (Hong *et al.*, 2014), was addressed in Papers III and IV.

3.2.5 Additional traits assessed

The following traits were addressed also in Paper II; cold hardiness (Ct) was estimated according to Nilsson & Walfridsson (1995) at ages 4, 9, 11 and 16 in the FS trial (F485) and at age three in the F726 OP trial; flower production (FP) and number of cones per tree (CO) were estimated at age 18 in the FS trial (F485). Finally, the number of seeds per cone (nSC) and the weight of 1000 seeds in grams (W), which are two traits routinely measured in the analysis and processing of seeds, were also considered.

3.3 Genotypic data

The marker data used for Paper II were previously described in Li *et al.* (2014). Briefly, a set of 492 FS individuals from trial F485 were genotyped with amplified polymorphism (AFLP) and SNP markers. After sorting and mapping simultaneously all marker data, 155 AFLP markers genotyped on 455 individuals and 166 SNP markers genotyped on 91 individuals were retained in the analysis; and two data sets were used for further analysis, the A dataset with AFLP only, and the S+A dataset with both SNP and AFLP data.

In Papers III and IV, 694 progeny trees and 46 parents were genotyped. DNA from vegetative buds was extracted and three GBS libraries were prepared and sequence on an Illumina HiSeq 2000 platform (at SciLifeLab, Sweden). SNP filtering, calling and a first baseline imputation of missing genotypes through LD K-nearest neighbour method (Money *et al.*, 2015) in TASSEL (Bradbury *et*

al., 2007) were performed commonly for Papers III and IV. In Paper III two additional and separate imputation methods were used to compare them in the subsequent analysis, i.e., random (RND) and expectation maximization algorithm (EM) imputation methods, under the same assumptions of minor allele frequency (MAF) cut-off of 1% and a missing data threshold of 10%; 8707 and 8719 SNP were retained respectively for RND and EM methods. After the baseline imputation, RND imputation was perform in Paper IV, assuming a MAF of 1% and a missing data rate of 5%, such that 6344 SNP were retained. RND and EM imputations were implemented in synbreed (Wimmer *et al.*, 2012) and rrBLUP (Endelman, 2011) packages in R (R Core Team, 2016).

3.4 Statistical analysis

3.4.1 Prior adjustment of phenotypic data

Prior to any genetic analysis the phenotypic measurements were adjusted for within trial environmental effects for all trials used in this thesis. A post-blocking procedure (Ericsson, 1997) was used to take into account the large-scale environmental variation in trials addressed in Papers I, III and IV. Normal-score transformation (Gianola & Norton, 1981) was performed on Vt assessments to linearize the data addressed in Paper I.

Univariate single site spatial analysis was performed (when possible) for all trials and traits (Papers I – IV), with the objective to adjust the data for within-trial microenvironmental effects. Spatial adjustments were performed using the row and column coordinates of the trial, and by fitting a linear mixed model with the residual structure incorporating only the experimental design elements as factors; if the spatial distribution of residuals were non-random for any trait, a second model was fitted, such that the residual component was structured as first-order separable autoregressive model (Dutkowski *et al.*, 2006; Dutkowski *et al.*, 2002; Gilmour, 1997).

The adjusted values of the wood properties addressed in Papers III – IV (*i.e.*, MFA, MOEs, DEN and MOEd) were calculated by removing the variation of the experimental design features and post-block effects, since spatial adjustments were not possible due to unavailable data for the full trial.

3.4.2 General linear mixed model

To estimate the genetic variance-covariance components, linear mixed models were used in the studies underlying this thesis (Papers I – IV). The general linear mixed model used was

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}, \quad (1.)$$

with

$\text{var}(\boldsymbol{y}) = \mathbf{ZGZ'} + \mathbf{R} = \mathbf{V}$, $\text{var}(\boldsymbol{u}) = \mathbf{G}$, $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{R}$

where $\boldsymbol{y}$ is the vector of individual tree adjusted phenotypic observations with expectation $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \mathbf{V})$, $\boldsymbol{\beta}$ is the vector of fixed effects; $\boldsymbol{u}$ is the vector of random effects with expectation $\boldsymbol{u} \sim N(0, \mathbf{G})$, and $\boldsymbol{\varepsilon}$ is the vector of residuals with expectation $\boldsymbol{\varepsilon} \sim N(0, \mathbf{R})$. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the respective incidence matrices of $\boldsymbol{\beta}$ and $\mathbf{u}$. Estimates of the fixed (BLUE) and random (BLUP) effects were obtained by solving the Henderson mixed model equation:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}Z} & \mathbf{Z'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix} \quad (2.)$$

where $\mathbf{R}$ and $\mathbf{G}$ are the variance-covariance matrices of the residuals and random effects, respectively.

Univariate and multivariate multi-environment trial analysis

To estimate variances, covariances, breeding values and Type-B genetic correlations, UV MET analyses were used in Papers I – II. In addition, with the aim to estimate more precise variances, covariances, Type-A, Type-B and Type-AB genetic correlations, MV MET analyses were also used in Paper I.

For the UV MET analysis, the $\mathbf{G}$ matrix was modelled for both studies (Papers I-II) using FA covariance structures following Smith *et al.* (2001) and Cullis *et al.* (2010). In Paper II, three different trials were analysed, thus we considered that fitting just one factor (FA1) was adequate; whereas in Paper I, the four populations were evaluated separately and five trials per population were analysed, therefore we compared the convenience of using different factors (FA1 and FA2) as well as extended-factor-analytic structures with up to three factors (XFA1, XFA2 and XFA3). Besides, we also modelled $\mathbf{G}$ with an unstructured variance-covariance matrix parameterized as heterogeneous covariances (US) and as heterogeneous correlations (CORGH), in Paper I. For all models fitted in the UV MET analysis (Papers I-II), residuals were assumed to be independent, with expectation $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_{\boldsymbol{\varepsilon}}^2)$, where $\mathbf{I}$ is the identity matrix and $\sigma_{\boldsymbol{\varepsilon}}^2$ is the residual variance.

In the case of the MV MET analysis, each variable at each trial was considered as a separate trait (trait-assessment-trial). To evaluate the precision of different variance-covariance structures in MV MET analysis, **G** was modelled by fitting the same seven different variance-covariance structures defined above (*i.e.*, CORGH, US, FA1, FA2, XFA1,XFA2 and XFA3; in this case, the **R** matrix was modelled using a US structure model, but with covariances set to zero for traits measured in different trials.

3.4.3 Family, pedigree and genomic relationships

In Paper I family model was used since no pedigree information other than the seed-parent identity was available, i.e., the vector $\boldsymbol{u}$ from Eq. 1 and Eq. 2 has an expectation of ~N(0, σ_f^2), where σ_f^2 is the family variance.

ABLUP-A and ABLUP-AD models

Pedigree information existed for the data used in Papers II, III and IV therefore, an individual (animal) model was used by incorporating the pedigree-based additive numerator relationship matrix (**A**) to the model (ABLUP-A model) such that the $\boldsymbol{u}$ vector from Eq. 1 and 2 has an expectation ~N(0, $\mathbf{A}\sigma_a^2$) where σ_a^2 is the pedigree-based additive genetic variance. Univariate ABLUP-A models were used in Papers III-IV.

The pedigree-based dominance numerator genetic relationship matrix (**D**) was estimated according to Lynch and Walsh (1998) in Paper IV, i.e., ABLUP-AD model, such that the Eq. 1 was extended to include both additive and non-additive effects as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_a a} + \boldsymbol{Z_d d} + \boldsymbol{\varepsilon} \qquad (3.)$$

where $\boldsymbol{a}$ and $\boldsymbol{d}$, are the vectors of additive and dominance random effects, respectively, which follow a normal distribution with correspondingly expectations $\boldsymbol{a}$ ~ N(0, $\mathbf{A}\sigma_a^2$) and $\boldsymbol{d}$ ~ N(0,$\mathbf{D}\sigma_d^2$); $\sigma_d^2$is the pedigree-based dominance genetic variance, and $\boldsymbol{Z_a}$ and $\boldsymbol{Z_d}$ are the incidence matrix for $\boldsymbol{a}$ and $\boldsymbol{d}$ respectively.

GBLUP-A, GBLUP-AD and GBLUP-ADE

Genomic relationships based on the SNP marker information describe above were calculated for the data addressed in Papers III-IV. A genomic best linear unbiased predictor (GBLUP) method with only additive effects was used in both papers, whereas in Paper IV genomic-based dominance (GBLUP-AD) and epistatic (GBLUP-ADE) effects were examined additionally. GBLUP-A is

derived from ABLUP-A but differs in that the **A** matrix is substituted by a genomic-based realized relationship matrix ($\mathbf{G_A}$) which is calculated from the SNP marker data according to VanRaden (2008), such that the vector **u** in Eq. 1 and the vector **a** in Eq. 3, follow a normal distribution with expectation ~N(0,$\mathbf{G_A}\sigma_a^2$) where σ_a^2 is now the additive genetic variance but based on SNP marker information.

The genomic-based dominance relationship matrix ($\mathbf{G_D}$) was estimated following Vitezica *et al.* (2013) such that the vector $\boldsymbol{d}$ in Eq. 3 follows now a normal distribution with expectation ~ N(0,$\mathbf{D}\sigma_d^2$). The genomic-based matrices based on the first-order epistatic interaction were calculated by the Hadamard products of $\mathbf{G_A}$ and $\mathbf{G_D}$, as $\mathbf{G_{AA}}$=$\mathbf{G_A}$#$\mathbf{G_A}$ (additive by additive terms), $\mathbf{G_{AD}}$=$\mathbf{G_A}$#$\mathbf{G_D}$ (additive by dominance terms) and $\mathbf{G_{DD}}$= $\mathbf{G_D}$ # $\mathbf{G_D}$ (dominance by dominance terms), with the following model including all effects:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_a a} + \boldsymbol{Z_d d} + \boldsymbol{Z_{e1} e_{aa}} + \boldsymbol{Z_{e2} e_{ad}} + \boldsymbol{Z_{e3} e_{dd}} + \boldsymbol{\varepsilon} \quad (4.)$$

where, $\boldsymbol{Z_{e1}}$, $\boldsymbol{Z_{e2}}$ and $\boldsymbol{Z_{e3}}$ are the incidence matrices for $\boldsymbol{e_{aa}}$, $\boldsymbol{e_{ad}}$ and $\boldsymbol{e_{dd}}$ which are the vectors of additive by additive, additive by dominance and dominance by dominance epistatic effects; and these components are assumed to follow a normal distribution with expectations ~N(0,$\mathbf{G_{AA}}\sigma_{aa}^2$), ~N(0,$\mathbf{G_{AD}}\sigma_{ad}^2$) and ~N(0,$\mathbf{G_{DD}}\sigma_{dd}^2$), respectively; where σ_{aa}^2, σ_{ad}^2 and σ_{dd}^2 are the additive by additive, additive by dominance and dominance by dominance epistatic interaction variances.

Bayesian methods

With the aim to estimate and compare prediction abilities based on SNP marker effects, Bayesian ridge regression (BRR) and Bayesian LASSO (BL) were applied to the data addressed in Paper III, with the BGLR package in R (Perez & de los Campos, 2014).

BRR method assumes that the vector $\boldsymbol{u}$ from Eq. 1 follows a multivariate normal prior distribution with a common variance to all marker effects, i.e., ~N(0, $\boldsymbol{I_p}\sigma_m^2$), where p is the number of markers, σ_m^2 is the unknown genetic variance which is contributed by each marker and assigned as $\sim\chi^{-2}(df_m, S_m)$, where df_m are the degrees of freedom and S_m the scale parameter.

The BL method assumes that vector $\boldsymbol{u}$ from Eq. 1 follows a hierarchical prior distribution with $\boldsymbol{u}$ ~N(0, $\boldsymbol{T}\sigma_m^2$), where $\boldsymbol{T} = \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2)$. τ_j^2 is assigned as $\tau_j^2 \sim Exp(\lambda^2)$, $j = 1, \ldots, p$. λ^2 is assigned as $\lambda^2 \sim Gamma(r, \delta)$.

In both cases, the residual variance is assigned as $\sigma_\varepsilon^2 \sim \chi^{-2}(df_\varepsilon, S_\varepsilon)$, where df_ε is degrees of freedom and S_ε is the scale parameter for residual variance.

For the QTL mapping study (Paper II) the de-biased LASSO approach by Javanmard and Montanari (2014) and Li *et al.* (2017b) was applied. In brief, this

approach differs from traditional LASSO estimator in that it constructs an unbiased LASSO estimator which asymptotically follow a normal distribution, and aims to calculate the *p*-values for all markers in the study, instead of only the markers selected by the standard LASSO.

3.4.4 Accuracies and predictive abilities of breeding values

The accuracy of the predicted pedigree-based EBV at Paper II was calculated for each F_1 mother as $r = \sqrt{1 - (PEV/\sigma_b^2)}$, where *PEV* is the prediction error variance derived from the elements of the inverse of the coefficient matrix of the mixed model equations, and σ_b^2 is the across-site additive variance. Prediction accuracy in Paper III, was defined as the Pearson product-moment correlation between the cross-validated GEBVs and the pedigree-based EBVs, *r(GEBV,EBV)*; and the predictive ability (PA), in Papers III-IV as the Pearson product-moment correlation between the cross-validated GEBV or GEGV and, the adjusted phenotypes, *r(GEBV, **y**) or r(GEGV, **y**).* In Paper IV, based on the full dataset, goodness-of-fit was calculated by estimating the correlation between the adjusted phenotypes and the total predicted genetic ($r(\hat{G}_{full}, \hat{Y}_{full})$) or additive values ($r(\hat{A}_{full}, \hat{Y}_{full})$). The fitted lined plot of the standard error of the predictions (SEPs) was evaluated to assess the precision of the predicted BVs among all models.

3.4.5 Selection response of Genomic Selection

Compared to the traditional phenotypic selection (TPS), the relative efficiency of GS, RE and RE per year (RE/year) were estimated at Paper III according to Grattapaglia & Resende (2011) as

$$RE = \frac{r(GEBV_{GS}, EBV)}{r(EBV_{TPS}, EBV)} \quad (5.)$$

Whereas, in Paper IV, the expected response of GS (RGS) was calculated following Resende *et al.* (2017) as

$$RGS(\%) = \left(\frac{\overline{EGV_S} - \overline{EGV_0}}{\overline{EGV_0}}\right) \times 100 \quad (6.)$$

where, $\overline{EGV_S}$ is the average of the expected genetic values and $\overline{EGV_0}$ is the population average.
By reducing 50% the breeding cycle (shortening or omitting the times of progeny testing), RE and RGS per year were also estimated.

3.4.6 Statistical software

All statistical analyses were executed in ASReml 3.0 (Gilmour *et al.*, 2009), ASReml 4.0 (Gilmour *et al.*, 2015) or ASReml-R (Butler *et al.*, 2009), except those already indicated above.

4 Results and discussion

4.1 Multi-environment analyses and G × E

4.1.1 Type-B genetic correlations

Multi-environment trials allow to detect the different performance of genotypes at different locations in the presence of G × E, and the Type-B genetic correlations can be used to detect the rank changes among several environments, as well as the degree of G × E (Burdon *et al.*, 2019; Ukrainetz *et al.*, 2018; Baltunis *et al.*, 2010; Burdon, 1977). Our results showed that all Type-B genetic correlations for the traits and sites studied in Paper II (see Table 3 in Paper II) were positive. However, they were very small for some traits and trials; indeed, small favourable Type-B genetic correlations were observed among trial F725 (Site 2 in Paper II) and the other two trials for Vt, whereas for DBH and Ht the smallest Type-B genetic correlations were detected among F726 (Site 3 in Paper II) and the remaining trials, showing the presence of a large G × E, at young ages. EBV through single site and through UV MET analysis (Figure 6), for DBH, Ht and Vt (at ages one and eight) could also be considered as sign of G × E, as they notably differ depending on the site.

Nevertheless, as results of Paper I showed, when UV MET analysis was used, genetic parameter estimations resulted in higher standard errors (SE), compared with MV MET analysis. Simulation studies in crops reported better accuracies and selection efficiencies, as well as estimates of genetic correlations when MV REML model was used (Viana *et al.*, 2010; Holland, 2006). Large SE were obtained in Paper II, but were not possible to compare with other methods, since only UV MET was used; still, we were able to compare the results from UV MET with those from MV MET analyses in Paper I, since in this study, both analyses were performed independently in four different unrelated populations.

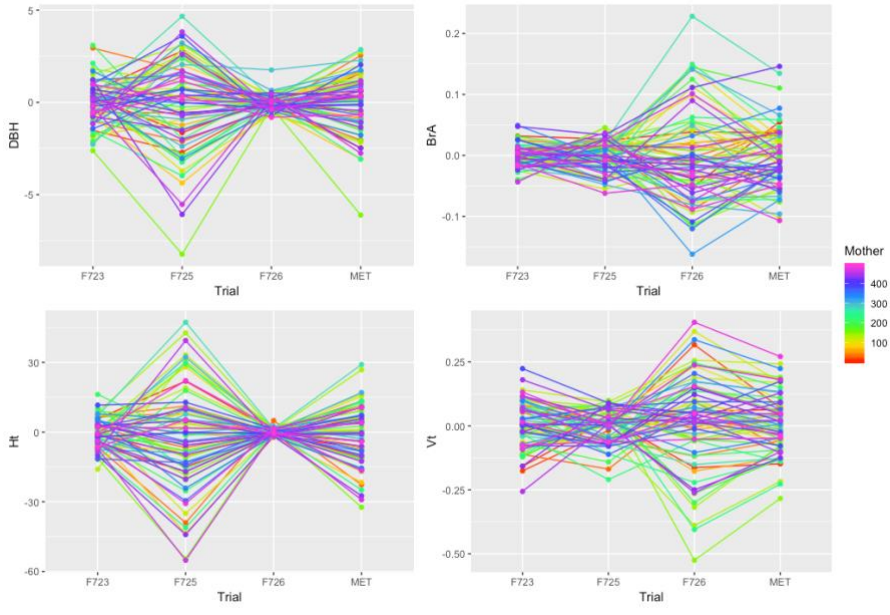

Figure 6. Estimated breeding values EBV for DBH, BrA, Ht and Vt. Overall EBV obtained through the univariate multi-environment analysis are represented by the acronym MET.

All SE for Type-B genetic correlations, in all trial series were generally higher when calculated from UV MET analysis than MV MET analysis (see Tables 6 and S2 from Paper I). XFA3 was the model that showed the lowest SE for both methods, and therefore the results presented here and in Paper I for the MV MET analysis, are based on the XFA3 model. Our results agree with previous studies in which it was reported that FA models are more effective to capture genetic variances and covariances and therefore, more accurate predictions of genotypes can be obtained (Walker *et al.*, 2019; Chen *et al.*, 2018a; Ogut *et al.*, 2014).

Based on the Type-B genetic correlations estimated from MV MET analysis, remarkable G × E was observed, at age 10 for both traits within trial series 2 and 3, among trials with the lowest Tsum and those with the largest Tsum; further, a similar pattern was discernible at age 20 for both traits. A trend was clearly recognisable by the moderate negative Pearson product-moment correlation (-0.41, $p < 0.001$), that showed how the Type-B genetic correlations for Ht at both ages, became smaller as the difference in Tsum between trials increased (Figure 7A).

The existence of G × E for growth traits has been previously reported for species such as *Pinus radiata* D. Don (Ivkovic *et al.*, 2015) or Norway spruce,

Picea abies L. (Chen *et al.*, 2017). Studies in *Pinus elliottii* Engelm. (Dieters *et al.*, 1995) and *Pinus taeda* L. (Xiang *et al.*, 2003), reported an increase on Type-B genetic correlation with the age, which was also observed for Scots pine in this thesis (Paper I), still it was possible to detect G × E for Ht and Vt at older ages.

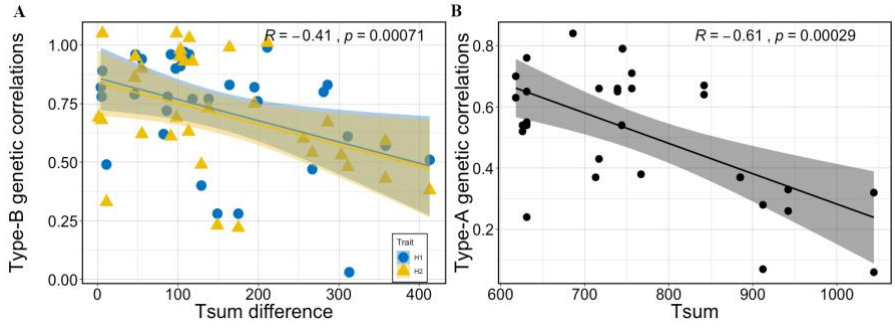

Figure 7. Plots of A) Type-B genetic correlations between Ht at different assessments against Tsum differences among trials, and B) Type-A genetic correlations among Vt and Ht against Tsum.

4.1.2 Type-A and Type-AB genetic correlations

Type-A and Type-AB genetic correlations were possible to detect through MV MET analysis implemented in Paper I (see Table 5 and Fig. 2 in Paper I). In trial series 2 and 3, Type-A genetic correlations between Vt and Ht were higher as the mortality and harshness (lower Tsum) of the trials increased, while it decreased as the trials became milder (higher Tsum); this pattern was observed by the strong negative Pearson product-moment correlation (-0.61, $p < 0.001$), among Tsum and Type-A genetic correlations between Vt and Ht, illustrated in Figure 7B, and that shows how the Type-A genetic correlations decreased as the Tsum of the trials increased. The sign of Type-AB genetic correlations between Vt and Ht, changed from positive to negative, as the differences in Tsum increased, which indicate that tree height is more dependent on the health of the trees in harsh environments. Additionally, it was observed that those trials with positive Type-A correlation were the same that showed significant genetic variation in susceptibility to *Phacidium infestans* in a previous study (Persson *et al.*, 2010). This fact in conjunction with the trend observed for Type-A and Type-B genetic correlations, indicates that height in harsh environments reflects the health of the trees, as Persson (2006) suggested.

No clear G × E was detected for trial series 1 and 4, with high positive Type-B genetic correlations, nevertheless both series represented more homogenous Tsum among most of the trials within series.

4.1.3 Multi-environment heritabilities and coefficients of variation

Within-trial narrow-sense estimated heritabilities ($\hat{h}_w^2$) were low to moderate for all traits addressed in Paper II, and they differ notably among sites (see Table 2b in Paper II); overall across-site narrow-sense heritabilities ($\hat{h}_b^2$) were generally smaller, which is also an indication of G × E. Most of the values estimated for narrow-sense heritabilities were similar to those reported in the literature for growth traits (DBH, Ht), BrA or Ct (Pagliarini *et al.*, 2016; Bian *et al.*, 2014; Prada *et al.*, 2014; Abrahamsson *et al.*, 2012). However, they were estimated through UV MET analysis and could be overestimated, as it was observed in results from Paper I, for which narrow-sense heritabilities (and SE) from UV MET analysis were higher than those form MV MET analysis, particularly for height (Figure 8). The $\hat{h}_w^2$ and the additive coefficient of variation ($\widehat{CV}_A$) for Ht increased with the age, consistent with previous studies in Scots pine in Sweden (Kroon *et al.*, 2011; Jansson *et al.*, 2003), but these parameters did not change for Vt. Besides, the environmental coefficients of variation ($\widehat{CV}_E$) increased as well for Ht in harsh trials and decreased at milder ones. On the contrary, $\widehat{CV}_A$ and $\widehat{CV}_E$ for Vt decreased with the age in harsh sites and increased in milder ones. The magnitudes of $\widehat{CV}_A$ and $\widehat{CV}_E$ estimates, agree with previous ones reported in Scots pine and other pine species (Baltunis & Brawner, 2010; Persson & Andersson, 2003; Haapanen, 2001).

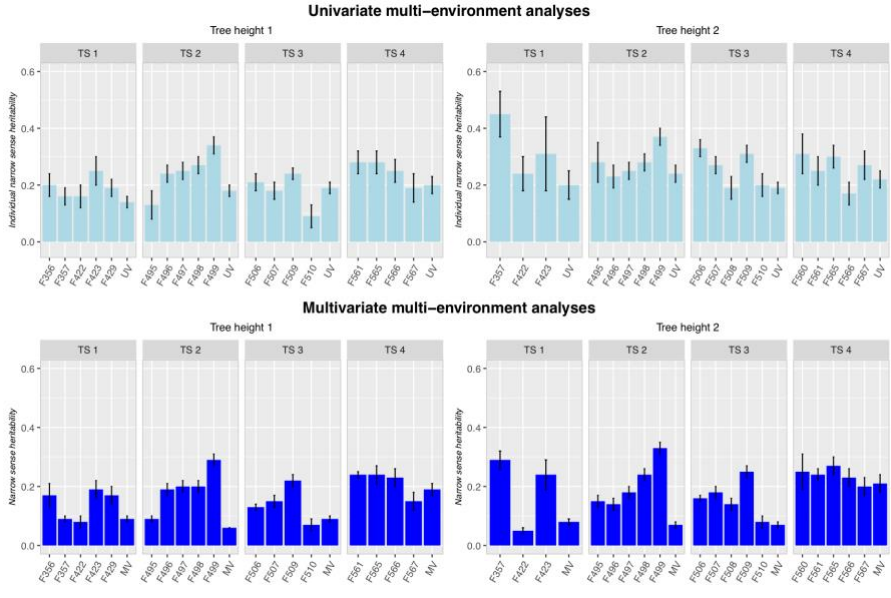

Figure 8. Within- and overall across-trial narrow-sense heritabilities for height estimated at 10 (tree height 1) and 20 (tree height 2) years, calculated in Paper I.

4.2 Genetic architecture of complex traits

The experiments of QTL studies in trees are usually based on phenotypic values from only one generation, normally a FS family with environmental and genetic factors confounded (Hall *et al.*, 2016; Thavamanikumar *et al.*, 2013; Lerceteau *et al.*, 2001). Although the candidate gene approach is not going to play an important role in the tree breeding, it is still important to decipher the relationship between genes and complex traits (Bartholome *et al.*, 2016a; Isik, 2014; Markussen *et al.*, 2003). In Paper II, 18 AFLP and 12 SNP markers were identified to be associated with QTL with medium to large effects, for one or more phenotypic traits, intercepts or EBV (see Tables 4 and D1, in Paper II); it was observed that when a QTL was associated simultaneously to a phenotype- and EBV-based QTL, the latter one always expressed a higher proportion of the variance. Besides, some QTL were detected for the same trait across several ages, making them possible candidates for early selection. QTL detection was not consistent across environments, probably due to the detected G × E or QTL × E, as it has been previously stated for different tree species (Bartholome *et al.*, 2013; Freeman *et al.*, 2013; Rae *et al.*, 2008; Groover *et al.*, 1994).

4.3 Genomic prediction

Nowadays GS is a subject undergoing intense study in animal, crop and tree breeding, as theoretical results seems to indicate that GS could be the tool that can aid to improve accuracy of breeding value predictions, reduce breeding cycles, increase selection intensities and obtain greater genetic gains per unit time (Grattapaglia *et al.*, 2018; Crossa *et al.*, 2017; Isik *et al.*, 2015; Grattapaglia, 2014; Isik, 2014). The GS studies underlying this thesis are the initial analyses to have a first indication of the possible advantages of genomic predictions for Scots pine in Sweden and are the basis for future GS studies in the specie (Papers III-IV).

4.3.1 Effect of the imputation method on predictions (Paper III)

For species as Scots pine, GBS is a good genotyping alternative to perform GS or GWAS studies, however this GBS data generate large amounts of missing sequences, thus filtering and imputation of SNPs are extremely important steps that could have an effect on predictions (Dodds *et al.*, 2015). Accordingly with previous studies that used GBS data (El-Dien *et al.*, 2018; El-Dien *et al.*, 2015; Poland *et al.*, 2012), generally the EM algorithm performed better than the RND imputation method, since slightly higher PAs and prediction accuracies for all genomic prediction models and traits were achieved (Table 2).

4.3.2 Impact of the model on predictions (Paper III)

The ABLUP model produced the highest prediction accuracies for all traits, among all models; however in terms of PA, genomic prediction models (GBLUP, BRR and BL) performed better for almost all traits analysed (Table 2). None of the genomic prediction models presented better predictions for all traits. Similar results were reported in other species where all genomic predictions models have similar impacts on PAs or prediction accuracies (Chen *et al.*, 2018b; Tan *et al.*, 2017; Thistlethwaite *et al.*, 2017; Bartholome *et al.*, 2016b; Isik *et al.*, 2016; Ratcliffe *et al.*, 2015). Among all genomic models and based on results and computational requirements, GBLUP was the most effective method in Paper III.

Table 2. *Predictive ability (PA) and prediction accuracy (Accuracy) of each model for each trait.*

Model	Type	Traits							
		Ht1	Ht2	DBH1	DBH2	MFA	MOEs	DEN	MOEd
ABLUP	PA	0.20	0.38	0.26	0.23	0.30	0.39	0.41	0.44
	Accuracy	0.83	0.81	0.83	0.84	0.83	0.75	0.81	0.82
GBLUP-EM	PA	0.20	0.39	**0.26**	0.26	0.29	0.39	**0.40**	0.41
	Accuracy	0.69	0.75	**0.73**	0.74	0.73	0.69	**0.73**	0.74
GBLUP-RND	PA	0.19	0.38	0.25	0.25	0.28	0.37	0.38	0.40
	Accuracy	0.67	0.74	0.71	0.72	0.71	0.67	0.71	0.72
BL-EM	PA	0.15	0.39	0.22	**0.30**	**0.33**	0.36	0.32	0.40
	Accuracy	0.66	0.74	0.70	**0.75**	**0.76**	0.67	0.69	0.71
BL-RND	PA	**0.26**	0.36	0.26	0.26	0.28	0.34	**0.40**	0.41
	Accuracy	0.69	0.73	0.71	0.72	0.68	0.65	0.71	0.72
BRR-EM	PA	0.18	**0.41**	0.25	0.27	**0.33**	**0.42**	**0.40**	**0.46**
	Accuracy	0.65	**0.77**	0.72	**0.75**	0.73	**0.70**	0.72	**0.76**
BRR-RND	PA	0.24	0.39	0.21	0.24	0.27	0.40	**0.40**	0.45
	Accuracy	**0.72**	0.75	0.70	0.74	0.73	0.68	0.72	0.75

4.3.3 Effect of TS, VS and number of SNPs (Paper III)

Based on the cross-validation results presented above, EM combined with GBLUP or BRR showed slightly better results among the genomic prediction models, and ABLUP in terms of accuracy, therefore were the methods choose to evaluate the effect of TS and VS on the predictions (Paper III). Agreeing to previous reports (Chen *et al.*, 2018b; Lenz *et al.*, 2017), all models had

increasing and similar patterns of PA for all traits, with the increasing sizes of TS (Figure 9A), whereas ABLUP presented the best accuracies for all traits and TS ratios. Using about 70–80% of individuals sampled in this population would produce similar PAs and accuracies as the full sample size.

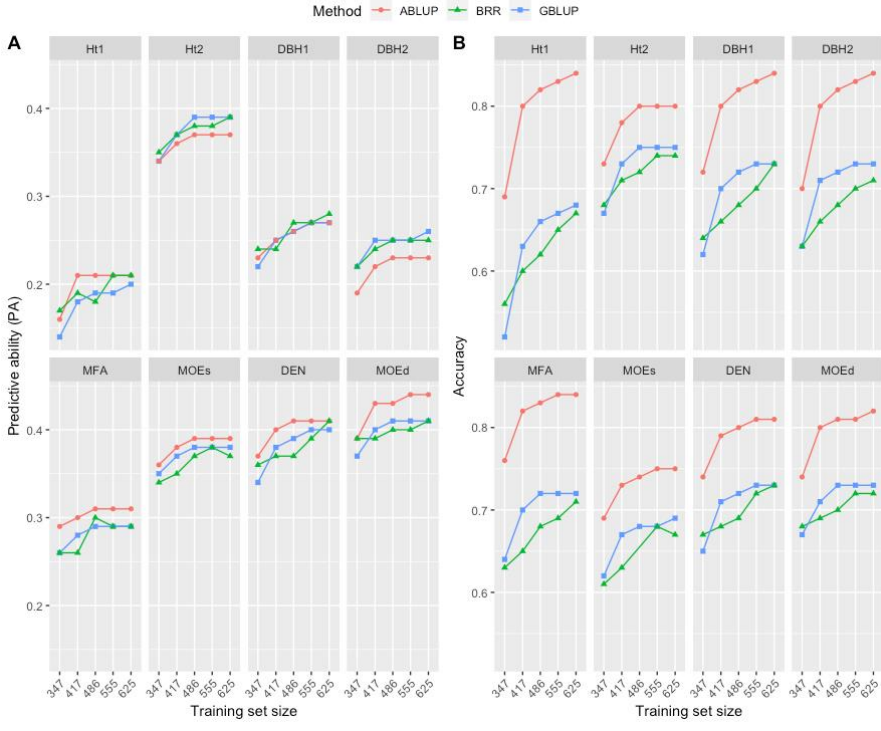

Figure 9. A) Predictive ability and prediction accuracy of ABLUP, BRR and GBLUP models for different sizes of training and validation sets.

The BRR method, combined with EM imputation, was used to evaluate the impact of number of SNPs on the predictions (see Fig. 2 in Paper III). It has been earlier reported that accuracies increased with the number of markers reaching a plateau when the number of markers are between 4K–8K (Chen *et al.*, 2018b; Lenz *et al.*, 2017; Tan *et al.*, 2017), but in the Scots pine population study in Paper III a plateau was not reached; however, only 3K – 4K SNPs were necessary to get similar efficiencies to those achieved for all 8719 SNPs for all traits.

4.3.4 Outcomes of non-additive effect on genomic predictions (Paper IV)

The use of genome-wide markers and GS has the additional advantage of decomposing the genetic variance into its additive and non-additive components without the need to perform specific mating designs to obtain FS families or replicated clonal material. In Paper IV, pedigree and genomic data were used to construct BLUP models that accounted for additive and non-additive (dominance and first order epistatic effects).

As expected, the cross-validated PA was slightly higher for all traits with ABLUP models regardless of the non-additive effects (Figure 10), yet very similar for both genomic and pedigree models, in concordance with studies in eucalyptus, Norway spruce and *Pinus taeda* (Chen *et al.*, 2018a; Bouvet *et al.*, 2016; de Almeida *et al.*, 2016).

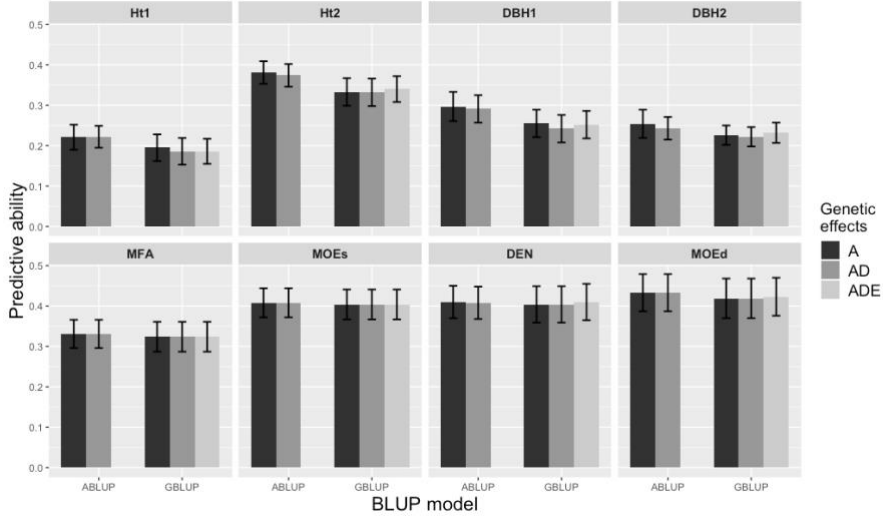

Figure 10. Mean of predictive abilities and standard errors for the different models and genetic effects.

Based on the full dataset and correlations between adjusted phenotypes and total or additive genetic values, additive values based on ABLUP fitted better than additive values with GBLUP, only for Ht30 and MOEd (Table 3). Tan *et al.* (2018) observed that most genetic values based on ABLUP-AD correlated better with phenotypes. In Paper IV, the highest correlations were detected among genetic values and adjusted phenotypes (0.76 – 0.97) for the GBLUP model with epistatic effects, which is in accordance to Bouvet *et al.* (2016) which found better fitting of GBLUP models. These results indicate that GBLUP epistatic model outperformed in the genetic response estimations the ABLUP and

GBLUP dominance models, and specially for some traits can end up in better estimations of genetic values.

Table 3. *Goodness-of-fit: correlation between adjusted phenotypes and additive- ($\hat{A}_{full}$) or total- ($\hat{G}_{full}$) genetic value of the full dataset. Correlations statistically significant at 0.01.*

Trait	Genetic effects	GBLUP $r(\hat{A}_{full},\hat{Y}_{full})$	GBLUP $r(\hat{G}_{full},\hat{Y}_{full})$	ABLUP $r(\hat{A}_{full},\hat{Y}_{full})$	ABLUP $r(\hat{G}_{full},\hat{Y}_{full})$
Ht1	A	0.72	.	0.68	.
	AD	0.72	0.72	0.68	**0.94**
	ADE	0.72	**0.83**	.	.
Ht2	A	0.79	.	0.84	.
	AD	0.79	0.81	0.84	**0.92**
	ADE	0.78	**0.97**	.	.
DBH1	A	0.73	.	0.74	.
	AD	0.73	0.73	0.74	0.74
	ADE	0.73	**0.76**	.	.
DBH2	A	0.72	.	0.71	.
	AD	0.72	0.77	0.70	**0.89**
	ADE	0.71	**0.95**	.	.
MFA	A	0.81	.	0.78	.
	AD	0.81	0.81	0.78	0.78
	ADE	0.81	0.81	.	.
MOEs	A	0.86	.	0.85	.
	AD	0.86	0.86	0.85	0.85
	ADE	0.86	0.86	.	.
DEN	A	0.86	.	0.87	.
	AD	0.86	0.86	0.87	0.87
	ADE	0.85	**0.95**	.	.
MOEd	A	0.85	.	0.88	.
	AD	0.85	0.85	0.88	0.88
	ADE	0.84	**0.97**	.	.

Comparable SEPs for most traits were seen regardless the model used, with slightly higher SEPs for GBLUP compared with ABLUP for Ht10 and MFA. GBLUP epistatic model exhibited the slight highest SEPs for almost all traits (see Fig.2, Paper IV), in contrast with studies in interior spruce (El-Dien *et al.*,

2018), white spruce (El-Dien *et al.*, 2016) or *Pinus taeda* (Munoz *et al.*, 2014) that showed the lowest SEPs when GBLUP epistatic model was used.

4.3.5 Genomic-based variance components and heritabilities.

Narrow-sense heritabilities were generally higher with ABLUP than GBLUP models (Papers III – IV), similar to other studies in conifers (Chen *et al.*, 2018b; El-Dien *et al.*, 2018; Lenz *et al.*, 2017); which can be explained by the genomic marker relationship matrix that consider the variation (deviation) among individuals of a family from the family average relatedness (Hayes *et al.*, 2009; VanRaden, 2008; White *et al.*, 2007).

Accounting for dominance effects in both ABLUP and GBLUP models, resulted in detection of dominance variance only for growth traits (see Table 1 in Paper IV). When dominance variance was identified, a diminution in both additive and residual variances was seen (Figure 11), however the proportion of dominance variance detected with ABLUP was higher than with GBLUP. For all growth traits and two wood quality traits, additional detection of additive by additive epistatic effects in GBLUP models, caused a further reduction of the additive and residual variance components, and as consequence, narrow-sense heritabilities were reduced and broad-sense heritabilities increased.

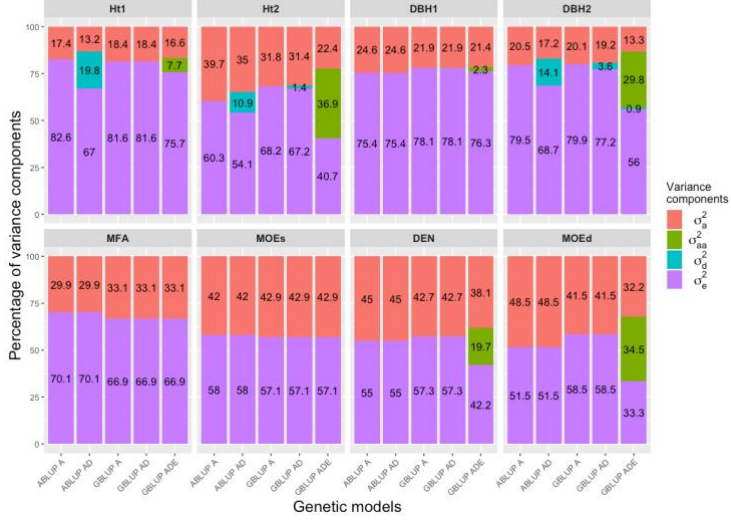

Figure 11. Percentages of the different variance components for each model and trait. $\hat{o}_a^2$, $\hat{o}_d^2$, $\hat{o}_e^2$ and $\hat{o}_{aa}^2$ denote estimated additive, dominance, residual and epistatic additive × additive variances.

As it has been stated in previous studies (El-Dien *et al.*, 2018; Tan *et al.*, 2018; Bouvet *et al.*, 2016; El-Dien *et al.*, 2016; Munoz *et al.*, 2014; White *et al.*, 2007), additive and non-additive effects are not independent; an inadequate estimation of genetic effects, and therefore overestimation of narrow-sense heritabilities, could be a consequence of the insufficient power of pedigree-based methods to discern between both effects.

In Scots pine, no clear patterns were noticed between prediction accuracies and narrow-sense heritabilities (Paper III), as it was previously stated for maritime pine and Norway spruce (Chen *et al.*, 2018b; Bartholome *et al.*, 2016b); however, a linear trend was found among PAs and narrow-sense heritabilities ($r = 0.99$, $p < 0.001$), as the traits with lowest narrow-sense heritabilities exhibited the lowest PAs and the highest PAs were detected for the traits with highest heritabilities (Figure 12); similar patterns were observed in other tree species such as loblolly pine (Resende *et al.*, 2012c) or maritime pine (Isik *et al.*, 2016).

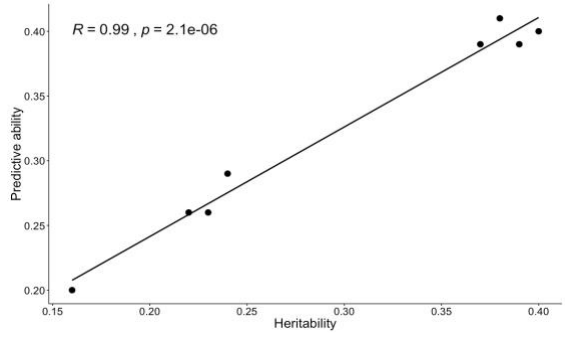

Figure 12. Regression plot among PAs and narrow-sense heritabilities based on GBLUP-EM model.

4.3.6 Relative GS efficiency and response of GS (Papers III – IV)

The Scots pine breeding cycle takes up to 21 or 36 years depending on the selection strategies used. One of the advantages of GS is the possibility to reduce the field testing periods and consequently the cycle length.

If the breeding cycle is reduced by 50% and 75% (omitting or reducing progeny test periods), the RE/year can doubled and triple, respectively (Ratcliffe *et al.*, 2015; Resende *et al.*, 2012b; Grattapaglia & Resende, 2011); in Scots pine, if the breeding cycle is shorten by 50%, the RE/year increased between 59 – 85% for GBLUP, 50 – 90% for BRR and 52 – 83% for BL (Table 4).

Table 4. *Relative efficiency (RE) and relative efficiency per year (RE/year) of genomic prediction models compared to ABLUP from cross validated models and for each trait.*

Trait	RE			RE[a]/year			RE[b]/year		
	GBLUP	BRR	BL	GBLUP	BRR	BL	GBLUP	BRR	BL
Ht1	0.83	0.78	0.8	1.66	1.57	1.59	1.59	1.5	1.52
Ht2	0.93	0.95	0.91	1.85	1.9	1.83	1.77	1.81	1.74
DBH1	0.88	0.87	0.84	1.76	1.73	1.69	1.68	1.66	1.61
DBH2	0.88	0.89	0.89	1.76	1.79	1.79	1.68	1.7	1.7
MFA	0.88	0.88	0.92	1.76	1.76	1.83	1.68	1.68	1.75
MOEs	0.92	0.93	0.89	1.84	1.87	1.79	1.76	1.78	1.71
DEN	0.9	0.89	0.85	1.8	1.78	1.7	1.72	1.7	1.63
MOEd	0.9	0.93	0.87	1.8	1.85	1.73	1.72	1.77	1.65

[a] and [b] represent first and second group of selection strategies, respectively.

RGS/year was estimated for GBLUP-AD and GBLUP-ADE using ABLUP-AD as a benchmark for the RPS/year, and again reducing the breeding cycle by 50%. Relative higher RGS/year, i.e., higher genetic gains were obtained with GBLUP models compared with ABLUP, especially remarkable for wood traits for which expected gains varied between 33 – 117%, and between 7 – 70% for growth traits, for the top 50 (7%) individuals (Figure 13); our results are in accordance with similar studies in Norway spruce (Chen *et al.*, 2018a), in which the genetic gains obtained from GBLUP with non-additive effects were higher for wood traits than growth traits, however in both cases superior than traditional phenotypic selection, whereas in eucalypts the greatest gains were observed for growth traits (Resende *et al.*, 2017).

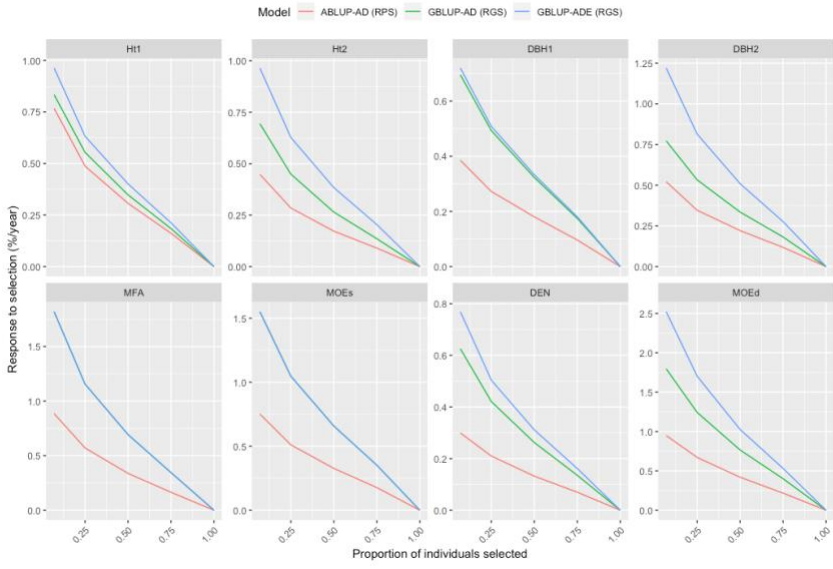

Figure 13. Expected response of genomic selection (RGS) as a percentage gain of the average population EGV per year of GBLUP-AD and GBLUP-ADE compared with the response of phenotypic selection (RPS) per year of ABLUP-AD, calculated for all traits and different proportions of individuals selected (7, 25, 50 and 100%).

5 Conclusions and future perspectives

In the studies underlying this thesis, we have combined new methods to the traditional quantitative genetics theory and applied for the first time (1) quantitative genomics, (2) factor-analytic methods and (3) multiple generation QTL study in Scots pine, to improve and understand genetic parameter estimations of complex traits of interest within the Swedish breeding program; and from which could be drawn the conclusions and future perspectives shown below.

Genotype-by-environment interactions were found for adaptive and growth traits using univariate multi-environment analysis (Papers I – II); however, the estimates of single variables may be overestimated as it was observed in the multivariate multi-environment analysis implemented with factor analytic structures, that allowed the study of up to 19 traits simultaneously, and resulted in more accurate estimations of genetic parameters (Paper I). Therefore, factor-analytic approach should be applied routinely in analysis of multiple genetically connected trials within the breeding program.

In harsh environments, trees were still affected by environmental perturbations at older ages, and the main cause of the G × E for tree vitality and height at young and older stages was the difference in temperature sum between trials (Paper I). A high positive genetic association between tree vitality and height on harsh sites was observed, but it decreased as the temperature sum increased, which suggests that tree growth on harsh and mild environments should be treated as separate traits and targeted to different deployment regions. Further studies are recommended for genetically connected populations, covering trials located at wider ranges of temperature sums, to confirm the patterns found in this thesis.

More accurate estimations were observed by using EBV than phenotypic data for the detection of QTL, since the former ones explained a higher percentage of the variance (Paper II); nevertheless, multivariate multi-environment analysis are recommended instead of univariate multi-environment to obtain more

precise EBV. Besides, within this thesis the number of phenotypic data was much larger than the number of EBV, which made that most of the QTL were detected based on the phenotypic data. Hence, further studies with analogous number of EBV and phenotypes are suggested to deepen in the patterns showed here and possibly to detect more candidate genes for early selection within the breeding program.

The genomic prediction studies underlying this thesis provide an initial perspective of the use of genomic data into prediction models in Scots pine. Growth and wood genetic parameters were more precisely estimated with genomic- than with pedigree-based models (Papers III – IV); and due to computational and predictive efficiency GBLUP was the most effective method to perform genomic predictions (Paper III). The high efficiency of GS in Scots pine encourages to develop GS strategies for the species' operational breeding program.

The incorporation of the non-additive effects to the genomic predictions, showed that by accounting for these effects, the estimations of additive and residual variances and associated heritabilities would be more accurate, and consequently more precise estimated breeding values and higher genetic gains can be achieved (Paper IV). To confirm whether the epistatic variation observed in this thesis plays and important role for growth and wood quality traits of Scots pine populations, a bigger population than the current studied here, should be genotyped and phenotyped.

While GBS is a good alternative for Scots pine as it was demonstrated in this thesis (Papers III – IV), faster and more reliable genotyping platforms should be developed. This is underway with current effort in de novo genome sequencing and re-sequencing in order to develop a SNP array for the Swedish Scots pine breeding populations.

One of the greatest advantages of GS is the possibility to reduce the breeding cycle; here, it was proven that by shortening the breeding cycle in 50% (assuming a reduction in progeny tests periods), the expected response of GS per year could be 50% – 90% higher than the traditional pedigree-based selection, depending on the selection strategy used (Papers III – IV).

Despite the promising results showed in this thesis, the population used to perform genomic predictions, was a small population with a shallow pedigree. Therefore, to understand better the utility and power of genomic selection for complex traits in Scots pine, more studies should be carried out using different populations, preferably with deeper pedigrees, tested through several generations, and at different environments, as well as, single-stage and multivariate genomic prediction approaches.

References

Abrahamsson, S., Nilsson, J.E., Wu, H., Gil, M.R.G. & Andersson, B. (2012). Inheritance of height growth and autumn cold hardiness based on two generations of full-sib and half-sib families of *Pinus sylvestris*. *Scandinavian Journal of Forest Research,* 27(5), pp. 405-413.

Andersson, B., Elfving, B., Ericsson, T., Persson, T. & Gregorsson, B. (2003). Performance of improved *Pinus sylvestris* in northern Sweden. *Scandinavian Journal of Forest Research,* 18(3), pp. 199-206.

Andersson, B., Elfving, B., Persson, T., Ericsson, T. & Kroon, J. (2007). Characteristics and development of improved *Pinus sylvestris* in northern Sweden. *Canadian Journal of Forest Research,* 37(1), pp. 84-92.

Andersson, B., Karlson, B., Ståhl, P. & Lindgre, D. (2011). Breeding objectives under changing environmental and market conditions. In: Rosvall, O. (ed.), *Review of the Swedish Tree Breeding Programme,*. Uppsala, Sweden: Skogforsk, pp. 15-19.

Baltunis, B.S. & Brawner, J.T. (2010). Clonal stability in *Pinus radiata* across New Zealand and Australia. I. Growth and form traits. *New Forests,* 40(3), pp. 305-322.

Baltunis, B.S., Gapare, W.J. & Wu, H.X. (2010). Genetic parameters and genotype by environment interaction in radiata pine for growth and wood quality traits in Australia. *Silvae Genetica,* 59(2-3), pp. 113-124.

Bärring, L., Berlin, M. & Andersson Gull, B. (2017). Tailored climate indices for climate-proofing operational forestry applications in Sweden and Finland. *International Journal of Climatology,* 37(1), pp. 123-142.

Bartholome, J., Bink, M.C., van Heerwaarden, J., Chancerel, E., Boury, C., Lesur, I., Isik, F., Bouffier, L. & Plomion, C. (2016a). Linkage and association mapping for two major traits used in the maritime pine breeding program: height growth and stem straightness. *PloS ONE,* 11(11), p. e0165323.

Bartholome, J., Salmon, F., Vigneron, P., Bouvet, J.M., Plomion, C. & Gion, J.M. (2013). Plasticity of primary and secondary growth dynamics in *Eucalyptus* hybrids: a quantitative genetics and QTL mapping perspective. *Bmc Plant Biology,* 13.

Bartholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C. & Bouffier, L. (2016b). Performance of genomic prediction within and across generations in maritime pine. *Bmc Genomics,* 17.

Bauer, A.M. & Leon, J. (2008). Multiple-trait breeding values for parental selection in self-pollinating crops. *Theoretical and Applied Genetics*, 116(2), pp. 235-242.

Beaulieu, J., Doerksen, T.K., MacKay, J., Rainville, A. & Bousquet, J. (2014). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *Bmc Genomics*, 15.

Beeck, C.P., Cowling, W.A., Smith, A.B. & Cullis, B.R. (2010). Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. *Genome*, 53(11), pp. 992-1001.

Berlin, M., Jansson, G. & Hogberg, K.-A. (2015). Genotype by environment interaction in the southern Swedish breeding population of *Picea abies* using new climatic indices. *Scandinavian Journal of Forest Research*, 30(2), pp. 112-121.

Bian, L.M., Shi, J.S., Zheng, R.H., Chen, J.H. & Wu, H.X. (2014). Genetic parameters and genotype-environment interactions of Chinese fir (*Cunninghamia lanceolata*) in Fujian Province. *Canadian Journal of Forest Research*, 44(6), pp. 582-592.

Bian, L.M., Zheng, R.H., Su, S.D., Lin, H.Z., Xiao, H., Wu, H.X. & Shi, J.S. (2017). Spatial analysis increases efficiency of progeny testing of Chinese fir. *Journal of Forestry Research*, 28(3), pp. 445-452.

Boratyński, A. (1991). Range of natural distribution. In: Giertych, M. and Matyas, C.(eds.), *Developments in Plant Genetics and Breeding*., Elsevier, pp. 19-30.

Bouvet, J.M., Makouanzi, G., Cros, D. & Vigneron, P. (2016). Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. *Heredity*, 116(2), pp. 146-157.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. & Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), pp. 2633-2635.

Burdon, R.D. (1977). Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genetica*, 26(5-6), pp. 168-175.

Burdon, R.D. & Li, Y. (2019). Genotype-environment interaction involving site differences in expression of genetic variation along with genotypic rank changes: simulations of economic significance, *Tree Genetics & Genomes*, 15(1), p. 2.

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J. (2009). ASReml-R reference manual. Department of Primary Industries & Fisheries, Brisbane.

Cappa, E.P., El-Kassaby, Y.A., Munoz, F., Garcia, M.N., Villalba, P.V., Klapste, J. & Poltri, S.N. (2017). Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. *Molecular Breeding*, 37(10). p. 125.

Caudullo, G., Welk, E. & San-Miguel-Ayanz, J. (2018). *Pinus sylvestris*.Chorological maps for the main European woody species. Data in Brief 12, pp. 662–666..

Chen, C., Mitchell, S.E., Elshire, R.J., Buckler, E.S. & El-Kassaby, Y.A. (2013). Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes*, 9(6), pp. 1537-1544.

Chen, Z.-Q., Baison, J., Pan, J., Westin, J., Garcia-Gil, M.R. & Wu, H.X. (2018a). Increased prediction ability in Norway spruce trials using a marker x environment interaction and non-additive genomic selection model. *bioRxiv*, p. 478404.

Chen, Z.Q., Baison, J., Pan, J., Karlsson, B., Andersson, B., Westin, J., Garcia-Gil, M.R. & Wu, H.X. (2018b). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *Bmc Genomics*, 19.

Chen, Z.Q., Helmersson, A., Westin, J., Karlsson, B. & Wu, H.X. (2018c). Efficiency of using spatial analysis for Norway spruce progeny tests in Sweden. *Annals of Forest Science*, 75(1).

Chen, Z.Q., Karlsson, B. & Wu, H.X. (2017). Patterns of additive genotype-by-environment interaction in tree height of Norway spruce in southern and central Sweden. *Tree Genetics & Genomes*, 13(1).

Crossa, J., Beyene, Y., Kassa, S., Perez, P., Hickey, J.M., Chen, C., de los Campos, G., Burgueno, J., Windhausen, V.S., Buckler, E., Jannink, J.L., Cruz, M.A.L. & Babu, R. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3-Genes Genomes Genetics*, 3(11), pp. 1903-1926.

Crossa, J., Perez-Rodriguez, P., Cuevas, J., Montesinos-Lopez, O., Jarquin, D., de Los Campos, G., Burgueno, J., Gonzalez-Camacho, J.M., Perez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J. & Varshney, R.K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science*. 22(11), pp. 961–975.

Cullis, B., Gogel, B., Verbyla, A. & Thompson, R. (1998). Spatial analysis of multi-environment early generation variety trials. *Biometrics*, 54(1), pp. 1-18.

Cullis, B.R., Jefferson, P., Thompson, R. & Smith, A.B. (2014). Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. *Theoretical and Applied Genetics*, 127(10), pp. 2193-2210.

Cullis, B.R., Smith, A.B., Beeck, C.P. & Cowling, W.A. (2010). Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. *Genome*, 53(11), pp. 1002-1016.

Danell, O., Wilhelmsson, L., Andersson Gull, B., Karlsson, B., Rosvall, O. & Wernern, M. (1993). *Breeding programmes in Sweden. Corrected reprint from: Lee, S.J. (ed), Progeny testing and breeding strategies.* (Proc. from a meeting with the Nordic group for tree breeding, Arbetsrapport nr 302 1995, Skogforsk). Edinburgh: Forestry Commission.

de Almeida, J.E.D., Guimaraes, J.F.R., Silva, F.F.E., de Resende, M.D.V., Munoz, P., Kirst, M. & Resende, M.F.R. (2016). The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity*, 117(1), pp. 33-41.

Dieters, M.J., White, T.L. & Hodge, G.R. (1995). Genetic parameter estimates for volume from full-sib tests of slash pine (*Pinus elliottii*). *Canadian Journal of Forest Research*, 25(8), pp. 1397-1408.

Dodds, K.G., McEwan, J.C., Brauning, R., Anderson, R.M., van Stijn, T.C., Kristjansson, T. & Clarke, S.M. (2015). Construction of relatedness matrices using genotyping-by-sequencing data. *Bmc Genomics*, 16.

Dutkowski, G., Ivkovic, M., Gapare, W.J. & Mcrae, T.A. (2016). Defining breeding and deployment regions for radiata pine in southern Australia. *New Forests*, 47(6), pp. 783-799.

Dutkowski, G.W., Silva, J.C.E., Gilmour, A.R. & Lopez, G.A. (2002). Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research*, 32(12), pp. 2201-2214.

Dutkowski, G.W., Silva, J.C.E., Gilmour, A.R., Wellendorf, H. & Aguiar, A. (2006). Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Canadian Journal of Forest Research*, 36(7), pp. 1851-1870.

Eiche, V. (1966). Cold damage and plant mortality in experimental provenance plantations with Scots pine in northern Sweden. Stud. For. Suec. 36, p. 219.

El-Dien, O.G., Ratcliffe, B., Klapste, J., Chen, C., Porth, I. & El-Kassaby, Y.A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *Bmc Genomics*, 16.

El-Dien, O.G., Ratcliffe, B., Klapste, J., Porth, I., Chen, C. & El-Kassaby, Y.A. (2016). Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. *G3-Genes Genomes Genetics*, 6(3), pp. 743-753.

El-Dien, O.G., Ratcliffe, B., Klapste, J., Porth, I., Chen, C. & El-Kassaby, Y.A. (2018). Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (*Picea glauca* x *engelmannii*). *Molecular Breeding*, 38(3).

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E. (2011). A Robust, Simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS ONE*, 6(5).

Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, 4(3), pp. 250-255.

Ericsson, T. (1997). Enhanced heritabilities and best linear unbiased predictors through appropriate blocking of progeny trials. *Canadian Journal of Forest Research*, 27(12), pp. 2097-2101.

Falconer, D. & Mackay, T.F.C. (1996). Introduction to quantitative genetics. Harlow, Essex, UK, Longmans Green, 3.

Freeman, G.H. (1973). Statistical-methods for analysis of genotype-environment interactions. *Heredity*, 31, pp. 339-354.

Freeman, J.S., Potts, B.M., Downes, G.M., Pilbeam, D., Thavamanikumar, S. & Vaillancourt, R.E. (2013). Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytologist*, 198(4), pp. 1121-1134.

Gezan, S.A., de Carvalho, M.P. & Sherrill, J. (2017). Statistical methods to explore genotype-by-environment interaction for loblolly pine clonal trials. *Tree Genetics & Genomes*, 13(1).

Gezan, S.A., Huber, D.A. & White, T.L. (2006). Post hoc blocking to improve heritability and precision of best linear unbiased genetic predictions. *Canadian Journal of Forest Research*, 36(9), pp. 2141-2147.

Gianola, D. & Norton, H.W. (1981). Scaling threshold characters. *Genetics*, 99(2), pp. 357-364.

Gilmour, A., Gogel, B., Cullis, B., Welham, S. & Thompson, R. (2015). ASReml user guide release 4.1 structural specification. VSN International Ltd, Hemmel Hempstead, UK.

Gilmour, A.R. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(3), pp. 269 - 293.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009). ASReml User Guide Release 3.0. VSN International Ltd, Hemmel Hempstead, UK.

Goddard, M.E. & Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6), pp. 381-391.

Gorjanc, G., Cleveland, M.A., Houston, R.D. & Hickey, J.M. (2015). Potential of genotyping-by-sequencing for genomic selection in livestock populations. *Genetics Selection Evolution*, 47.

Grattapaglia, D. (2014). Breeding forest trees by genomic selection: current progress and the way forward. In: *Genomics of Plant Genetic Resources*, Springer, Dordrecht, pp. 651-682.

Grattapaglia, D. & Resende, M.D.V. (2011). Genomic selection in forest tree breeding. *Tree Genetics & Genomes*, 7(2), pp. 241-255.

Grattapaglia, D., Silva-Junior, O.B., Resende, R.T., Cappa, E.P., Müller, B.S.F., Tan, B., Isik, F., Ratcliffe, B. & El-Kassaby, Y.A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science*, 9(1693).

Groover, A., Devey, M., Fiddler, T., Lee, J., Megraw, R., Mitchel-Olds, T., Sherman, B., Vujcic, S., Williams, C. and Neale, D., 1994. Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics*, *138*(4), pp.1293-1300..

Haapanen, M.J.F.g. (2001). Time trends in gentic parameter estimates and selection efficiency for Scots pine in relation to field testing method. *Forest Genetics*, 8(2), pp. 129-144.

Hall, D., Hallingbäck, H.R. & Wu, H.X. (2016). Estimation of number and size of QTL effects in forest tree traits. *Tree Genetics & Genomes*, 12(6).

Hannrup, B., Jansson, G. & Danell, O. (2008). Genotype by environment interaction in *Pinus sylvestris* L. in Southern Sweden. *Silvae Genetica*, 57(6), pp. 306-311.

Hardner, C.M., Dieters, M., Dale, G., DeLacy, I. & Basford, K.E. (2010). Patterns of genotype-by-environment interaction in diameter at breast height at age 3 for eucalypt hybrid clones grown for reafforestation of lands affected by salinity. *Tree Genetics & Genomes*, 6(6), pp. 833-851.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K. & Goddard, M.E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, 41.

He, X.L. & Zhang, J.Z. (2006). Toward a molecular understanding of pleiotropy. *Genetics*, 173(4), pp. 1885-1891.

Holland, J.B. (2006). Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. *Crop Science*, 46(2), pp. 642-654.

Hong, Z., Fries, A. & Wu, H.X. (2014). High negative genetic correlations between growth traits and wood properties suggest incorporating multiple traits selection including economic weights for the future Scots pine breeding programs. *Annals of Forest Science*, 71(4), pp. 463-472.

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests*, 45(3), pp. 379-401.

Isik, F., Bartholome, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C. & Bouffier, L. (2016). Genomic selection in maritime pine. *Plant Science*, 242, pp. 108-119.

Isik, F., Holland, J. & Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding*. Springer International Publishing, New York.

Isik, F., Kumar, S., Martinez-Garcia, P.J., Iwata, H. & Yamamoto, T. (2015). Acceleration of forest and fruit tree domestication by genomic selection. *Advances in Botanical Research,* 74, pp. 93-124.

Isik, F., Li, B.L. & Frampton, J. (2003). Estimates of additive, dominance and epistatic genetic variances from a clonally replicated test of loblolly pine. *Forest Science,* 49(1), pp. 77-88.

Isik, F., Li, B.L., Frampton, J. & Goldfarb, B. (2004). Efficiency of seedlings and rooted cuttings for testing and selection in *Pinus taeda*. *Forest Science,* 50(1), pp. 44-53.

Ivkovic, M., Gapare, W., Yang, H.X., Dutkowski, G., Buxton, P. & Wu, H. (2015). Pattern of genotype by environment interaction for radiata pine in southern Australia. *Annals of Forest Science,* 72(3).

Jansson, G., Li, B.L. & Hannrup, B. (2003). Time trends in genetic parameters for height and optimal age for parental selection in Scots pine. *Forest Science,* 49(5), pp. 696-705.

Javanmard, A. & Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research,* 15, pp. 2869-2909.

Kelly, A.M., Cullis, B.R., Gilmour, A.R., Eccleston, J.A. & Thompson, R. (2009). Estimation in a multiplicative mixed model involving a genetic relationship matrix. *Genetics Selection Evolution,* 41.

Kelly, A.M., Smith, A.B., Eccleston, J.A. & Cullis, B.R. (2007). The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science,* 47(3), pp. 1063-1070.

Krakau, U.-K., Liesebach, M., Aronen, T., Lelu-Walter, M.-A. & Schneck, V. (2013). Scots Pine (Pinus sylvestris L.). In: Pâques, L.E. (ed.), *Forest Tree Breeding in Europe: Current State-of-the-Art and Perspectives.* Dordrecht: Springer Netherlands, pp. 267-323. https://doi.org/10.1007/978-94-007-6146-9_6

Kroon, J., Ericsson, T., Jansson, G. & Andersson, B. (2011). Patterns of genetic parameters for height in field genetic tests of *Picea abies* and *Pinus sylvestris* in Sweden. *Tree Genetics & Genomes,* 7(6), pp. 1099-1111.

Lenz, P.R.N., Beaulieu, J., Mansfield, S.D., Clement, S., Desponts, M. & Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *Bmc Genomics,* 18.

Lerceteau, E., Szmidt, A.E. & Andersson, B. (2001). Detection of quantitative trait loci in *Pinus sylvestris* L. across years. *Euphytica,* 121(2), pp. 117-122.

Li, Y.J., Suontama, M., Burdon, R.D. & Dungey, H.S. (2017). Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. *Tree Genetics & Genomes,* 13(3).

Li, Z., Guo, B., Yang, J., Herczeg, G., Gonda, A., Balazs, G., Shikano, T., Calboli, F.C.F. & Merila, J. (2017). Deciphering the genomic architecture of the stickleback brain with a novel multilocus gene-mapping approach. *Molecular Ecology,* 26(6), pp. 1557-1575.

Li, Z.T., Hallingback, H.R., Abrahamsson, S., Fries, A., Gull, B.A., Sillanpaa, M.J. & Garcia-Gil, M.R. (2014). Functional multi-locus QTL mapping of temporal trends in Scots pine wood traits. *G3-Genes Genomes Genetics,* 4(12), pp. 2365-2379.

Lind, M., Kallman, T., Chen, J., Ma, X.F., Bousquet, J., Morgante, M., Zaina, G., Karlsson, B., Elfstrand, M., Lascoux, M. & Stenlid, J. (2014). A *Picea abies* linkage map based on SNP

markers identifies QTLs for four aspects of resistance to *Heterobasidion parviporum* infection. *PloS ONE*, 9(7).

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Mäkinen, H. & Colin, F. (1998). Predicting branch angle and branch diameter of Scots pine from usual tree measurements and stand structural information. *Canadian Journal of Forest Research*, 28(11), pp. 1686-1696.

Markussen, T., Fladung, M., Achere, V., Favre, J.M., Faivre-Rampant, P., Aragones, A., Perez, D.D., Harvengt, L., Espinel, S. & Ritter, E. (2003). Identification of QTLs controlling growth, chemical and physical wood property traits in *Pinus pinaster* (Ait.). *Silvae Genetica*, 52(1), pp. 8-15.

Mathew, B., Holand, A.M., Koistinen, P., Leon, J. & Sillanpaa, M.J. (2016). Reparametrization-based estimation of genetic parameters in multi-trait animal model using Integrated Nested Laplace Approximation. *Theoretical and Applied Genetics*, 129(2), pp. 215-225.

Mátyás, C., Ackzell, L. & Samuel, C. (2004). *EUFORGEN technical guidelines for genetic conservation and use for Scots pine (Pinus sylvestris)*: Bioversity International.

Meuwissen, T., Hayes, B. & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6(1), pp. 6-14.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), pp. 1819-1829.

Meyer, K. (2009). Factor-analytic models for genotype x environment type problems and structured covariance matrices. *Genetics Selection Evolution*, 41.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.Y. & Myles, S. (2015). LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3-Genes Genomes Genetics*, 5(11), pp. 2383-2390.

Mullin, T.J., Andersson, B., Bastien, J., Beaulieu, J., Burdon, R., Dvorak, W., King, J., Kondo, T., Krakowski, J. & Lee, S. (2011). Economic importance, breeding objectives and achievements. In: Plomion, C. & Bousquet, J. (eds), *Genetics, Genomics and Breeding of Conifers*, pp. 40-127.

Munoz, P.R., Resende, M.F.R., Gezan, S.A., Resende, M.D.V., de los Campos, G., Kirst, M., Huber, D. & Peter, G.F. (2014). Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics*, 198(4), pp. 1759-+.

Neale, D.B. & Williams, C.G. (1991). Restriction-fragment-length-polymorphism mapping in conifers and applications to forest genetics and tree improvement. *Canadian Journal of Forest Research*, 21(5), pp. 545-554.

Nilsson, J.E. & Walfridsson, E.A. (1995). Phenological variation among plus-tree clones of *Pinus-sylvestris* (L) in northern Sweden. *Silvae Genetica*, 44(1), pp. 20-28.

Ogut, F., Maltecca, C., Whetten, R., McKeand, S. & Isik, F. (2014). Genetic analysis of diallel progeny test data using factor analytic linear mixed models. *Forest Science*, 60(1), pp. 119-127.

Pagliarini, M.K., Kieras, W.S., Moreira, J.P., Sousa, V.A., Shimizu, J.Y., Moraes, M.L.T., Furlani, E. & Aguiar, A.V. (2016). Adaptability, stability, productivity and genetic parameters in slash pine second-generation families in early age. *Silvae Genetica*, 65(1), pp. 71-82.

Pâques, L.E. (2013). Introduction. In: Pâques, L.E. (ed.), *Forest Tree Breeding in Europe: Current State-of-the-Art and Perspectives*. Dordrecht: Springer Netherlands, pp. 1-9. https://doi.org/10.1007/978-94-007-6146-9_1.

Perez, P. & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics,* 198(2), pp. 483-U63.

Persson, B. & Ståhl, E. (1993). Effects of provenance transfer and spacing in an experimental series of Scots pine (*Pinus sylvestris* L.) in northern Sweden. *Rapport-Sveriges Lantbruksuniversitet, Institutionen foer Skogsproduktion (Sweden)*.

Persson, T. (2006). *Genetic expression of Scots pine growth and survival in varying environment*. Dissertation. Swedish University of Agricultural Sciences.

Persson, T. & Andersson, B. (2003). Genetic variance and covariance patterns of growth and survival in northern *Pinus sylvestris*. *Scandinavian Journal of Forest Research,* 18(4), pp. 332-343.

Persson, T. & Andersson, B. (2004). Accuracy of single- and multiple-trait REML evaluation of data including non-random missing records. *Silvae Genetica,* 53(3), pp. 135-139.

Persson, T., Andersson, B. & Ericsson, T. (2010). Relationship between Autumn Cold Hardiness and Field Performance in northern *Pinus sylvestris*. *Silva Fennica,* 44(2), pp. 255-266.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S.Y., Manes, Y., Dreisigacker, S., Crossa, J., Sanchez-Villeda, H., Sorrells, M. & Jannink, J.L. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *Plant Genome,* 5(3), pp. 103-113.

Poland, J.A. & Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome,* 5(3), pp. 92-102.

Prada, E., Alia, R., Climent, J. & Diaz, R. (2014). Seasonal cold hardiness in maritime pine assessed by different methods. *Tree Genetics & Genomes,* 10(3), pp. 689-701.

R Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria.

Rae, A.M., Pinel, M.P.C., Bastien, C., Sabatti, M., Street, N.R., Tucker, J., Dixon, C., Marron, N., Dillen, S.Y. & Taylor, G. (2008). QTL for yield in bioenergy *Populus*: identifying GxE interactions from growth at three contrasting sites. *Tree Genetics & Genomes,* 4(1), pp. 97-112.

Ratcliffe, B., El-Dien, O.G., Cappa, E.P., Porth, I., Klapste, J., Chen, C. & El-Kassaby, Y.A. (2017). Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. *G3-Genes Genomes Genetics,* 7(3), pp. 935-942.

Ratcliffe, B., El-Dien, O.G., Klapste, J., Porth, I., Chen, C., Jaquish, B. & El-Kassaby, Y.A. (2015). A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* x *glauca*) using unordered SNP imputation methods. *Heredity,* 115(6), pp. 547-555.

Resende, M.D.V., Resende, M.F.R., Sansaloni, C.P., Petroli, C.D., Missiaggia, A.A., Aguiar, A.M., Abad, J.M., Takahashi, E.K., Rosado, A.M., Faria, D.A., Pappas, G.J., Kilian, A. & Grattapaglia, D. (2012a). Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist,* 194(1), pp. 116-128.

Resende, M.F.R., Munoz, P., Acosta, J.J., Peter, G.F., Davis, J.M., Grattapaglia, D., Resende, M.D.V. & Kirst, M. (2012b). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist,* 193(3), pp. 617-624.

Resende, M.F.R., Munoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F. & Kirst, M. (2012c). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics,* 190(4), pp. 1503-+.

Resende, R.T., Marcatti, G.E., Pinto, D.S., Takahashi, E.K., Cruz, C.D. & Resende, M.D.V. (2016). Intra-genotypic competition of *Eucalyptus* clones generated by environmental heterogeneity can optimize productivity in forest stands. *Forest Ecology and Management,* 380, pp. 50-58.

Resende, R.T., Resende, M.D.V., Silva, F.F., Azevedo, C.F., Takahashi, E.K., Silva, O.B. & Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity,* 119(4), pp. 245-255.

Rosvall, O. (2011). Review of the Swedish tree breeding program. *Skogforsk, Uppsala, Sweden*.

Rosvall, O., Jansson, G., Andersson, B., Ericsson, T., Karlsson, B., Sonesson, J. & Stener, L. (2001). Genetic gain from present and future seed orchards and clone mixes. *Redogorelse SkogForsk,* 1, p. 41.

Ruotsalainen, S. & Persson, T. (2013). Scots pine–*Pinus sylvestris* L. In: Mullin, T.J. & Lee, S.J. (eds.), *Best Practice for Tree Breeding in Europe. Skogforsk, Uppsala*, pp. 49-64.

Silva, J.C.E., Potts, B.M., Bijma, P., Kerr, R.J. & Pilbeam, D.J. (2013). Genetic control of interactions among individuals: contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. *New Phytologist,* 197(2), pp. 631-641.

Smith, A., Cullis, B. & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics,* 57(4), pp. 1138-1147.

Smith, A.B. & Cullis, B.R. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica,* 214(8).

Smith, A.B., Ganesalingam, A., Kuchel, H. & Cullis, B.R. (2015). Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theoretical and Applied Genetics,* 128(1), pp. 55-72.

Stefansson, E. & Sinko, M.J.S.F.S. (1967). Experiments with provenances of Scots pine with special regard to high-lying forests in northern Sweden, *Studia Forestalia Suecica,* 47, p. 108.

Strauss, S.H., Lande, R. & Namkoong, G. (1992). Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research,* 22(7), pp. 1050-1061.

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., McKinley, R. & Dungey, H. (2018). Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity*. 122(3), p. 370.

Tan, B., Grattapaglia, D., Wu, H.X. & Ingvarsson, P.K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Science,* 267, pp. 84-93.

Tan, B.Y., Grattapaglia, D., Martins, G.S., Ferreira, K.Z., Sundberg, B. & Ingvarsson, P.K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F-1 hybrids. *Bmc Plant Biology,* 17.

Thavamanikumar, S., Southerton, S.G., Bossinger, G. & Thumma, B.R. (2013). Dissection of complex traits in forest trees - opportunities for marker-assisted selection. *Tree Genetics & Genomes,* 9(3), pp. 627-639.

Thistlethwaite, F.R., Ratcliffe, B., Klapste, J., Porth, I., Chen, C., Stoehr, M.U. & El-Kassaby, Y.A. (2017). Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *Bmc Genomics*, 18.

Ukrainetz, N.K., Yanchuk, A.D. & Mansfield, S.D. (2018). Climatic drivers of genotype-environment interactions in lodgepole pine based on multi-environment trial data and a factor analytic model of additive covariance. *Canadian Journal of Forest Research*, 48(7), pp. 835-854.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), pp. 4414-4423.

Viana, J.M.S., Sobreira, F.M., de Resende, M.D.V. & Faria, V.R. (2010). Multi-trait BLUP in half-sib selection of annual crops. *Plant Breeding*, 129(6), pp. 599-604.

Vitezica, Z.G., Varona, L. & Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, 195(4), pp. 1223-1230.

Walker, T.D., Isik, F. & McKeand, S.E. (2019). Genetic Variation in Acoustic Time of Flight and Drill Resistance of Juvenile Wood in a Large Loblolly Pine Breeding Population. *Forest Science*.

Walsh, B. & Lynch, M. (2018). *Evolution and Selection of Quantitative Traits*: Oxford University Press.

White, T.L., Adams, W.T. & Neale, D.B. (2007). *Forest Genetics*: Cabi.

Wilhelmsson, L. & Andersson, B. (1993). Breeding of Scots pine (*Pinus sylvestris*) and lodgepole pine (*Pinus contorta* ssp. latifolia). *Corrected repring from: Progeny testing and breeding strategies, proceedings of the Nordic group of tree breeding. Edinburgh: Forestry Commission*, p. 184.

Wimmer, V., Albrecht, T., Auinger, H.J. & Schon, C.C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics*, 28(15), pp. 2086-2087.

Xiang, B., Li, B. & Isik, F. (2003). Time trend of genetic parameters in growth traits of *Pinus taeda* L. *Silvae Genetica*, 52(3-4), pp. 114-121.

Zas, R. (2006). Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. *Tree Genetics & Genomes*, 2(4), pp. 177-185.

Popular science summary

Scots pine is the second most important specie in Sweden, both from commercial and ecological points of view. Its wood and derivates are used among others, for construction, furniture, chip-boards, wood pulp, paper and energy products. To be able to determine if we can improve tree growth, survival and its properties, the tree characteristics and their variation, such as height, diameter, wood density, ability to survive or adapt, and resistance to pests and diseases, need to be understood.

Forest management and tree breeding aid to improve the profitability of forest products, through silvicultural, economical, genetical and statistical principles. Tree breeding aims to understand and decompose the genetic variation of trees for the characters of interest, so that the quantity that is genetically transmitted from one generation to another can be predicted; this allows an early selection of individuals with the desirable characteristics, so that there is no need to complete the long rotation periods for breeding selection. However, it is necessary to test the trees in the field and that they reach a certain age to be able to evaluate them, which make that the tree breeding cycle of Scots pine lasts between 20 to 30 years. For example, for Scots pine in harsh areas of northern Sweden, large mortality occurs predominantly during the first 12 to 16 years after planting, therefore evaluation of trees in such trials requires longer times than in southern Sweden.

The studies underlying this thesis aim to evaluate new statistical methodologies to improve the accuracy of the genetic variation estimates on growth, adaptive and wood quality traits.

Different families had been evaluated in several locations which allowed us to detect their different behaviour, and by using new statistical methods we were able to evaluate more characters simultaneously; this allowed us to observe that in harsh areas the trees are subject to a toughest environmental stress than in mild areas and for a longer time, and it seems to be related with the temperature.

In this thesis, a new methodology known as genomic selection was evaluated for the first time in Scots pine. One of the advantages of this methodology is that the trees can be evaluated at the seedling stage, which make that the tree improvement cycle can be shortened by omitting the evaluation of the trees in the field which, it can improve the profitability per year at higher rates than traditional methodology. By using this method we obtained more accurate estimation of the genetic variation of growth and wood quality traits, and we observed that by reducing the breeding cycle in 50%, genetic gain can increase between 50%–90%.

Populärvetenskaplig sammanfattning

Tall är Sveriges näst viktigaste trädslag, både ur kommersiella och ekologiska synvinklar. Dess trä och beståndsdelar används bland annat för konstruktion, möbler, spånskivor, vedmassa, papper och energiprodukter. För att kunna avgöra om vi kan förbättra trädtillväxt, överlevnad och dess egenskaper, måste trädets egenskaper och deras variation, såsom höjd, diameter, trädensitet, förmåga att överleva eller anpassa sig och motståndskraft mot skadedjur och sjukdomar förstås.

Skogsförvaltningen och skogsträdsförädlingen kan bidra till att förbättra lönsamheten för skogsprodukter genom skogsbruk, ekonomiska, genetiska och statistiska metoder. Skogsträdsförädlingen syftar till att förstå och särskilja den genetiska variationen för träd för de egenskaper som är av intresse, så att den mängd som är genetiskt överförd från en generation till en annan kan förutsägas; detta möjliggör tidigt urval av individer med önskvärda egenskaper, så att det inte finns något behov av att slutföra de långa rotationsperioderna för urval. Trots detta så är det nödvändigt att testa träden i fält, där de måste nå en viss ålder för att kunna utvärderas, vilket gör att förädlingscykeln för tall är mellan 20 till 30 år. För tall i kärva områden i norra Sverige så förekommer det hög dödlighet huvudsakligen under de första 12-16 åren efter plantering, därför krävs utvärdering av träd i sådana försök under lång tid än i södra Sverige.

De studier som ligger till grund för denna avhandling syftar till att utvärdera nya statistiska metoder för att förbättra noggrannheten i de genetiska variationskalkylerna på tillväxtegenskaper, adaptiva egenskaper och träkvalitetsegenskaper.

Olika familjer har utvärderats på flera olika lokaler som gjorde det möjligt för oss att upptäcka deras olika beteenden och genom att använda nya statistiska metoder kunde vi utvärdera flera egenskaper samtidigt. Detta gjorde det möjligt för oss att observera att i kärva områden är andelen träd som överlever mycket lägre än i milda områden. Träd i kärva områden utsätts för större miljöpåverkan

än i milda områden och under längre tid, och det verkar vara relaterat till temperatur.

I denna avhandling utvärderades en ny metod som kallas genomiskt urval för första gången i tall. En av fördelarna med den här metoden är att träden kan utvärderas som fröplanta, vilket gör att förädlingscykeln kan förkortas genom att utelämna utvärderingen av träden i fält, vilket kan förbättra lönsamheten per år i högre uppskattning än traditionell metodik. Genom att använda den här metoden fick vi en mer exakt uppskattning av den genetiska variationen av tillväxt och träkvalitetsegenskaper, och vi observerade att genom att minska förädlingscykeln med 50% kan genetisk vinst öka med 50% – 90%.

Acknowledgements

First of all, I would like to thank my main supervisor, **Harry Wu** and **Bengt Andersson Gull** for giving me the opportunity to do this project in one of the most attractive topics that may exist such as tree breeding. **Harry,** I want to express my gratitude for arranging the PhD courses, within the 2nd School of Forest Genetics, Biotechnology and Breeding, with some of the top international professors within their topics. And of course, for the wonderful trip to Brazil and USA which gave us the chance to visit some of the most advance tree breeding programs. I would like also to express my gratitude to my co-supervisors **Rosario García-Gil**, **Sara Abrahamsson**, **Fikret Isik** and **Pär Ingvarsson**.

Sara and **Sari** for your mentoring, your patience and finding always time to reply my emails and answer my questions. Thank you for proposing a plan B when I even didn´t know that I needed it. From the personal perspective I am very thankful for your support and understanding, for being there available for a conversation in the moments that was needed. By the way, we need a lunch meeting together, it was long ago when we had the last one.

Fikret for being always available for supervisors' meetings, fast mail replies, statistical advices, code checking and some car ridings in Raleigh. Your book was really helpful during this long path, it has become my sacred book and you have created a "monster" that loves factor analysis.

Torgny, jag är dig tacksam för att jag fick samarbeta i ditt projekt. I början var det verkligen svårt att förstå problemet med kärva och milda miljöer men i slutet var detta arbete brytningspunkten i min avhandling. Jag har utfört miljoner av analyser och min kunskap har avsevärt förbättrats genom det.

All my collaborators for their help, but a special mention to **Tim Mullin** and **Mikko Sillanpää** for your advices and interesting questions, and **Zhiqiang-Chen** for sharing your knowledge of R and ASReml.

Bengt, **Johan W** och **Sara**, tack så mycket for allt men speciellt för hjälpen och stödet i de svåraste stunderna.

Mari Suontama, thank you very much for your interest in my work, your advices and contribution.

Tack så mycket **allihopa i Skogforsk - Sävar**, för att få mig att trivas så bra där; **Anita** jag saknar redan dina historier. **Maggan** och **Kristina**, jag tycker att det är dags för annan afterwork.

Simon Birve, Ingela Sandström, Anders Eman, and all the technical staff at UPSC, thank you for all your help during this long 5 years "trip".

Within the research school I had the great fortune to meet some wonderful colleagues, **Johanna**, **Biyue** (you saved me with the ASReml-R installation, thank you), **Irena**, **Alexis** and **Christoffer**, **Julia** (I already miss you a lot); people used to tell me that Swedes do not talk too much and it's difficult to be their friend, but you **Jenny**, from the first minute, have proven that it's not like that.

Thank you to everyone at Harry´s and Sari´s labs, **Hong Zhou**, **Amaryllis Vidalis**, **David Hall**, **Henrik Hällingbäck**, **Lynn, John Baison**, **Jin Pan**, **Anders Fries**, **Sonali Ranade** (I am really happy that you are back in Umeå) and **Hong Nguyen** (I will never forget our time together at NCSU).

Rosario Sierra de Grado, the person who introduce me into the tree improvement world in the year 2008; muchas gracias por la oportunidad que me diste y ojalá algún día nuestros caminos se vuelvan a encontrar porque fue un placer aprender y trabajar a tu lado.

Johanna Witzell, who brought me to Sweden and gave me the opportunity to work in the wonderful world of Forest Pathology. It seems that with the climate change more and more pathogens are appearing so, who knows, maybe we can collaborate again in the near future.

And finally the most important people in my life, **my family**, that is what I miss the most of Asturias (gracies **Ali** y **Abri** por ayudame cola portada), pero especialmente quiero dar les gracies a los **mios padres**, por apoyame siempre, sobretó nos malos momentos que ye cuando fai falta. Y por supuestu a **René**, por quereme acompañar nesti viaxe, faes que'l camín seya munchu más afayadizu.